UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Microbiologia

Talita Bastos Machado

VÍRUS GIGANTES DE DNA:
sobre a genômica, proteômica da partícula e evolução de cedratvírus

Belo Horizonte
2025

Talita Bastos Machado

VÍRUS GIGANTES DE DNA:
sobre a genômica, proteômica da partícula e evolução de cedratvírus

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Microbiologia.

Orientador: Prof. Jônatas Santos Abrahão
Coorientadores: Dra. Ana Cláudia dos Santos Pereira Andrade e Prof. Luiz Eduardo Vieira Del Bem

Belo Horizonte
2025

043 Machado, Talita Bastos.
Vírus gigantes de DNA: sobre a genômica, proteômica da partícula e evolução de cedratvírus [manuscrito] / Talita Bastos Machado. – 2025.
170 f. : il. ; 29,5 cm.

Orientador: Prof. Jônatas Santos Abrahão. Coorientadores: Dra. Ana Cláudia dos Santos Pereira Andrade e Prof. Luiz Eduardo Vieira Del Bem.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-graduação em Microbiologia.

1. Microbiologia. 2. Vírus Gigantes. 3. Genoma. 4. Proteômica. I. Abrahão, Jônatas Santos. II. Andrade, Ana Cláudia dos Santos Pereira. III. Bem, Luiz Eduardo Vieira Del. IV. Universidade Federal e Minas Gerais. Instituto de Ciências Biológicas. V. Título.

CDU: 579

Ficha Catalográfica elaborada pela bibliotecária Fabiane C M Reis CRB 6 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PÓS-GRADUAÇÃO EM MICROBIOLOGIA

ATA DE DEFESA DE TESE

ATA DA DEFESA DE TESE DE **TALITA BASTOS MACHADO**

Nº REGISTRO: 2021695764

Às 14:00 horas do dia **10 de janeiro de 2025**, reuniu-se, por via remota, a Comissão Examinadora composta pelos Drs. Rodrigo Araújo Lima Rodrigues (Departamento de Microbiologia/ICB/UFMG), Juliana Reis Cortines (Universidade Federal do Rio de Janeiro), Savio Torres de Farias (Universidade Federal da Paraíba), Gabriel Magno de Freitas Almeida (The Artic University of Norway), o Prof. Dr. Jônatas Santos Abrahão (Orientador), a Dra. Ana Cláudia dos Santos Pereira Andrade (Coorientadora) e o Prof. Dr. Luiz Eduardo Vieira Del Bem (Coorientador), para julgar o trabalho final **"Vírus gigantes de DNA: sobre a genômica, proteômica da partícula e evolução de cedratvírus"** da aluna **Talita Bastos Machado**, requisito final para a obtenção do Grau de **DOUTORA EM CIÊNCIAS BIOLÓGICAS: MICROBIOLOGIA**. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Jônatas Santos Abrahão, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para a apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. A candidata foi considerada **APROVADA**. O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ata, que será assinada por todos os membros participantes da Comissão Examinadora. A candidata tem 60 (sessenta) dias, a partir desta data, para entregar a versão final da tese ao Programa de Pós-graduação em Microbiologia da UFMG e requerer seu diploma.

Belo Horizonte, 10 de janeiro de 2025

Membros da Banca:

Prof. Dr. Rodrigo Araújo Lima Rodrigues

Profa. Dra. Juliana Reis Cortines

Prof. Dr. Savio Torres de Farias

Prof. Dr. Gabriel Magno de Freitas Almeida

De acordo:

Prof. Dr. Jônatas Santos Abrahão

(Orientador)

Dra. Ana Cláudia dos Santos Pereira Andrade

(Coorientadora)

Prof. Dr. Luiz Eduardo Vieira Del Bem

(Coorientador)

Prof. Dr. Daniel de Assis Santos

(Coordenador do Programa de Pós-graduação em Microbiologia)

Documento assinado eletronicamente por **Savio Torres de Farias**, **Usuário Externo**, em 13/01/2025, às 10:49, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Gabriel Magno de Freitas Almeida**, **Usuário Externo**, em 13/01/2025, às 10:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Ana Cláudia dos Santos Pereira Andrade**, **Usuário Externo**, em 13/01/2025, às 11:01, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jonatas Santos Abrahao**, **Professor do Magistério Superior**, em 13/01/2025, às 13:09, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Rodrigo Araújo Lima Rodrigues**, **Professor do Magistério Superior**, em 13/01/2025, às 13:44, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Juliana Reis Cortines**, **Usuário Externo**, em 13/01/2025, às 16:24, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luiz Eduardo Vieira Del Bem**, **Chefe**, em 14/01/2025, às 10:09, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel de Assis Santos**, **Coordenador(a) de curso de pós-graduação**, em 23/01/2025, às 15:45, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3827351** e o código CRC **7742627E**.

Às minhas irmãs,
que fazem a vida valer a pena
e pelas quais eu morreria
sem pensar duas vezes.

AGRADECIMENTOS

Às minhas irmãs, Pâmela e Bianca, por serem as mulheres mais incríveis que conheço, por me ensinarem todos os dias o significado do amor. Por serem minha rocha nos momentos mais difíceis e por celebrarem comigo as minhas vitórias. Por serem minhas amigas além de tudo, e por acreditarem em mim muito mais do que eu mesma. Por tudo isso e muito mais eu dediquei o nome desse vírus a vocês (Pam + Bi = Pambiensis). Vocês são o que eu tenho de mais precioso na vida. Eu amo muito vocês.

À minha mãe, Silvana, por me ensinar a ser forte e guerreira, correr atrás dos meus objetivos e não desistir. Até hoje, quando eu coloco uma ideia na cabeça, eu vou atrás, da mesma forma que minha mãe fez a vida inteira.

À tia Ci, tio Joel e suas filhas, por terem ajudado durante muitos anos com a minha criação e das minhas irmãs e, com certeza, foram fundamentais nas nossas vidas. Eu não consigo expressar o tamanho da minha gratidão e do meu amor por cada um;

Aos meus de fé, Mari e Chico, que me suportam desde a graduação e nunca largaram minha mão. E muitas vezes me senti mais acolhida pelas famílias deles do que pela minha própria. Meu carinho por vocês é indescritível.

Aos meus queridos amigos Isabella, Bruna, Ana e Matheus que me aturaram muito no laboratório e fora dele, e fizeram meus dias muito mais leves e felizes. O que começou apenas como colegas de trabalho que eram do mesmo grupo de pesquisa se tornou uma amizade linda. Eu já não me vejo mais sem vocês, desejo carregá-los para sempre.

Ao professor Luís e todos os envolvidos no projeto NAVIO, pela oportunidade de participar desse projeto incrível e poder conhecer o Pantanal. Eu sou extremamente grata.

Ao Professor Frank Aylward por me aceitar no Virginia Tech e me proporcionar a oportunidade de fazer meu doutorado sanduíche em seu laboratório. Por ser também um excelente orientador, eu tenho apenas coisas boas a falar. Foi uma experiência incrível, tanto pessoal quanto profissional.

A todos os amigos que fiz no período do meu doutorado sanduíche que me ajudaram a me sentir acolhida. Em especial eu gostaria de agradecer ao Ali e a Paula que, definitivamente, me fizeram sentir muito bem-vinda. Além deles, eu agradeço imensamente à Lis, uma mineira que eu precisei sair do Brasil pra conhecer, e que sorte a minha. Vocês três foram os melhores presentes que os EUA me proporcionou.

Ao meu orientador Jônatas, por ter me orientado de uma forma impecável desde a graduação, por ser sempre presente e se preocupar com meu futuro profissional. Eu não poderia pedir por um orientador melhor;

Aos meus colegas do grupo GEPVIG, por todo o apoio, colaboração e companheirismo sempre. Me deixam orgulhosa em dizer que eu faço parte de um dos melhores grupos de pesquisa de vírus gigantes do Brasil e do mundo.

A todos os colegas do Laboratório de Vírus, que foi onde conheci pessoas incríveis que com certeza contribuíram demais para o meu crescimento pessoal e profissional.

A todos os professores do Lab Vírus, Erna, Giliane, Betânia e Rodrigo por enriquecer nosso conhecimento tanto em reuniões quanto em disciplinas;

À banca avaliadora pela disponibilidade e pela colaboração na fase final deste trabalho;

Às agências de fomento, CAPES, CNPq E FAPEMIG pelo auxílio financeiro que permitiu a realização deste trabalho;

Muito obrigada!

RESUMO

A descoberta dos vírus gigantes (VGs) em 2003 chamou muita atenção da comunidade científica principalmente devido ao grande tamanho de suas partículas e complexidade de seus genomas. Os genomas gigantescos dos VGs são uma das características mais intrigantes da virosfera. A transferência lateral de genes e até mesmo a criação de genes *de novo* têm sido propostas para explicar o gigantismo genômico dos VGs. No entanto, esses mecanismos parecem estar restritos a famílias específicas de VGs. Um fenômeno mais universal ainda não foi identificado. Aqui descrevemos a descoberta do cedratvirus pambiensis, um VG de ameba isolado no Brasil. Apesar de cdv. pambiensis ser muito semelhante a outros cedratvírus, o sequenciamento genômico revelou ser o maior genoma de cedratvírus já descrito, com 623.564 pares de bases e 842 proteínas preditas (entre elas, 76 ORFans). As razões desse maior genoma foram investigadas e revelaram um número sem precedentes de genes parálogos. Quase 73% do genoma de cdv. pambiensis é composto por genes com duas ou mais cópias. Grandes famílias/clusters de genes parálogos foram identificados, codificando diversas funções de proteínas preditas, e estão amplamente distribuídos no genoma. A investigação aprofundada sobre os mecanismos e origens da duplicação de genes revela que tanto a cópia em tandem proximal quanto a duplicação de segmento cromossômico contribuem para a expansão genômica de cdv. pambiensis. Finalmente, uma análise abrangente dos genomas de todos os grupos do domínio *Varidnaviria* conhecidos sugere que a duplicação de genes está relacionada ao gigantismo genômico em vários nucleocitovírus. A expansão de genomas virais por duplicações sucessivas seguidas por evolução independente de cada cópia do gene pode estar relacionada ao surgimento de novas funções gênicas e melhor adaptação viral a uma variedade de nichos. Na presente tese avançamos na caracterização do proteoma da partícula do cedratvírus. Nossas análises revelaram a presença de 283 proteínas virais, o correspondente a 33,6% das proteínas preditas no genoma, além de 172 proteínas associadas ao seu hospedeiro, *Acanthamoeba castellanii*. Análises comparativas entre os proteomas do cdv. pambiensis e do pithovirus sibericum, um vírus relacionado filogeneticamente, revelaram que 38,5% das proteínas identificadas são exclusivas para o cdv. pambiensis. Entretanto, 91 proteínas são compartilhadas por ambos os vírus, incluindo proteínas chaves envolvidas em diversos processos celulares, como replicação, transcrição e metabolismo. Todos esses achados têm nos ajudado a compreender diferentes aspectos biológicos sobre os VGs, além de elucidar aspectos filogenéticos e evolutivos sobre este fascinante grupo de vírus.

Palavras-chave: Vírus gigantes, cedratvírus, gigantismo genômico, paralogia, proteômica.

ABSTRACT

The discovery of giant viruses (GVs) in 2003 attracted much attention from the scientific community mainly due to the large size of their particles and the complexity of their genomes. The gigantic genomes of GVs are one of the most intriguing features of the virosphere. Lateral gene transfer and even de novo gene creation have been proposed to explain the genomic gigantism of GVs. However, these mechanisms seem to be restricted to specific GV families. A more universal phenomenon has not yet been identified. Here we describe the discovery of cedratvirus pambiensis, an amoeba GV isolated in Brazil. Although cdv. pambiensis is very similar to other cedratviruses, genome sequencing revealed it to be the largest cedratvirus genome ever described, with 623,564 base pairs and 842 predicted proteins (among them, 76 ORFans). The reasons for this larger genome were investigated and revealed an unprecedented number of paralogous genes. Almost 73% of the cdv. pambiensis is composed of genes with two or more copies. Large families/clusters of paralogous genes have been identified, encoding several predicted protein functions, and are widely distributed in the genome. In-depth investigation into the mechanisms and origins of gene duplication reveals that both proximal tandem copying and chromosome segment duplication contribute to the genomic expansion of cdv. pambiensis. Finally, a comprehensive analysis of the genomes of all known Varidnaviria groups suggests that gene duplication is related to genomic gigantism in several nucleocytoviruses. The expansion of viral genomes by successive duplications followed by independent evolution of each gene copy may be related to the emergence of new gene functions and improved viral adaptation to a variety of niches. In this thesis, we advanced in the characterization of the cedratvirus particle proteome. Our analyses revealed the presence of 283 viral proteins, corresponding to 33.6% of the predicted proteins in the genome, in addition to 172 proteins associated with its host, *Acanthamoeba castellanii*. Comparative analyses between the proteomes of cdv. pambiensis and pithovirus sibericum, a phylogenetically related virus, revealed that 38.5% of the identified proteins are exclusive to cdv. pambiensis. However, 91 proteins are shared by both viruses, including key proteins involved in several cellular processes, such as replication, transcription and metabolism. All these findings have helped us to understand different biological aspects of VGs, in addition to elucidating phylogenetic and evolutionary aspects of this fascinating group of viruses.

Keywords: Giant viruses, cedratvirus, genomic gigantism, paralogy, proteomics.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

APMV: Acanthamoeba polyphaga mimivírus

AVL: Amebas de vida livre

DNA: Ácido desoxirribonucleico

ICTV: International Committee on Taxonomy of Viruses

Kb: Kilobase

Mb: Megabase

MET: Microscopia eletrônica de transmissão

MEV: Microscopia eletrônica de varredura

NCLDV: Vírus grandes nucleocitoplasmáticos de DNA (do inglês "Nucleo-cytoplasmic large DNA viruses")

ORFs: Janelas abertas de leitura (do inglês, "*Open Reading Frames*")

PCR: Reação em cadeia da polimerase (do inglês "Polymerase chain reaction")

TGH: Transferência gênica horizontal

VG: Vírus gigantes

SUMÁRIO

CONTEÚDO **PÁG.**

1 - INTRODUÇÃO

1.1- O grupo dos vírus grandes nucleocitoplasmáticos de DNA (NCLDV) e sua importância para a virosfera

Até o início dos anos 2000, a maioria dos vírus era conhecida como entidades ultramicroscópicas e filtráveis em filtro de 0,2 µm, além de possuírem pequenos genomas que codificam poucas dezenas de proteínas (1). A descoberta de alguns vírus que infectam amebas, pertencentes ao grupo dos vírus grandes nucleocitoplasmáticos de DNA (NCLDV - do inglês Nucleo-cytoplasmic large DNA viruses) revelou genomas e partículas muito maiores que o descrito para os demais vírus (2,3). O grupo ganhou maior notoriedade com a descoberta do Acanthamoeba polyphaga mimivírus (APMV), um vírus que possui partículas com aproximadamente 750 nm e um genoma com 1,2 megabases (Mb) (4,5). Essas características destoam de alguns aspectos descritos para a maioria dos vírus conhecidos até então pois, é possível a visualização destes vírus por microscopia óptica, eles são retidos em filtro de 0,2 µm e seu genoma é maior que o genoma dos outros vírus já conhecidos e é até mesmo maior que os genomas de algumas bactérias como *Mycoplasma genitalium* (580 kb), *Ureaplasma urealyticum* (752 kb) *Buchnera sp.* (641 kb) e *Wigglesworthia brevipalpis* (698 kb) (4).

Os NCLDVs são ubíquos e infectam uma enorme gama de hospedeiros, incluindo algas, protistas, insetos, peixes, aves, répteis e mamíferos (2). Devido a essas características, este grupo vem se expandindo com o constante isolamento de novos vírus. Atualmente o filo *Nucleocytoviricota* está incluído no domínio *Varidnaviria*, reino *Bamfordvirae* e é composto por duas classes, cinco ordens e onze famílias (6), além de alguns vírus que ainda não foram alocados em nenhuma família. Neste contexto, os NCLDVs têm sido referidos como nucleocitovírus, em referência ao filo *Nucleocytoviricota*.

A virosfera possui uma história evolutiva polifilética. Não é conhecido nenhum gene presente em todos os vírus. Entretanto, dentro dos NCLDVs, existe um conjunto de genes chamado de “core genes”, que são compartilhados por quase todos os membros do grupo sugerindo, portanto, uma provável ancestralidade comum (7). Esses vírus codificam genes de proteínas necessárias para a transcrição, como DNA polimerases, helicases e topoisomerases, tornando esses vírus parcialmente independentes do sistema transcricional da célula hospedeira

(2,5,8). Além dos “core genes”, esses vírus compartilham outras características como partículas grandes ou gigantes (100 nm a 2300 nm), ciclo de multiplicação com fase citoplasmática, semelhança no arranjo estrutural de proteínas que compõem o capsídeo (double jelly roll) e genomas extensos (até 2,7 Mb). Todas essas características, juntas, indicam um perfil hipoteticamente monofilético para os nucleocitovírus (9,10).

Em 2010, Boyer e colaboradores propuseram uma análise dos domínios da vida através de análises filogenéticas envolvendo genes codificadores de proteínas relacionadas ao metabolismo de DNA e controle de transcrição, por exemplo, e não sequências de RNA ribossomal. Esta proposta foi justificada pelo fato de os genes de metabolismo estarem presentes nos genomas tanto de alguns vírus de DNA quanto nos genomas de organismos celulares. Dessa forma, após análises filogenéticas utilizando tais genes, foi sugerido que os NCLDVs comporiam um quarto domínio da vida, separado dos outros três já existentes (11).

O isolamento de novos vírus de amebas contribuiu para a expansão do conhecimento acerca dos NCLDVs, possibilitando a realização de novas análises, o que levou outros pesquisadores a refutar essa hipótese do quarto domínio da vida. Em 2014, Yutin e colaboradores realizaram uma nova reconstrução filogenética, incluindo novos grupos de vírus de amebas (Pandoravírus e pithovírus), suas análises demonstraram que, ao contrário do que preconizava a hipótese do quarto domínio, boa parte dos genes do “mundo celular” encontrados nos vírus gigantes foram adquiridos por transferência gênica horizontal através de seus hospedeiros eucarióticos. Portanto, os vírus gigantes não representariam um quarto domínio da vida que evoluiu a partir de uma simplificação de um organismo celular derivado do último ancestral comum universal (12). Esta segunda hipótese é a mais aceita atualmente (13,14,15). Nesta hipótese, os vírus gigantes provavelmente surgiram a partir de ancestrais de fagos do grupo dos tectivírus, que foram incorporados por células proto-eucarióticas durante o fenômeno de endogenização de bactérias, que posteriormente se tornaram as atuais mitocôndrias (Figura 1).

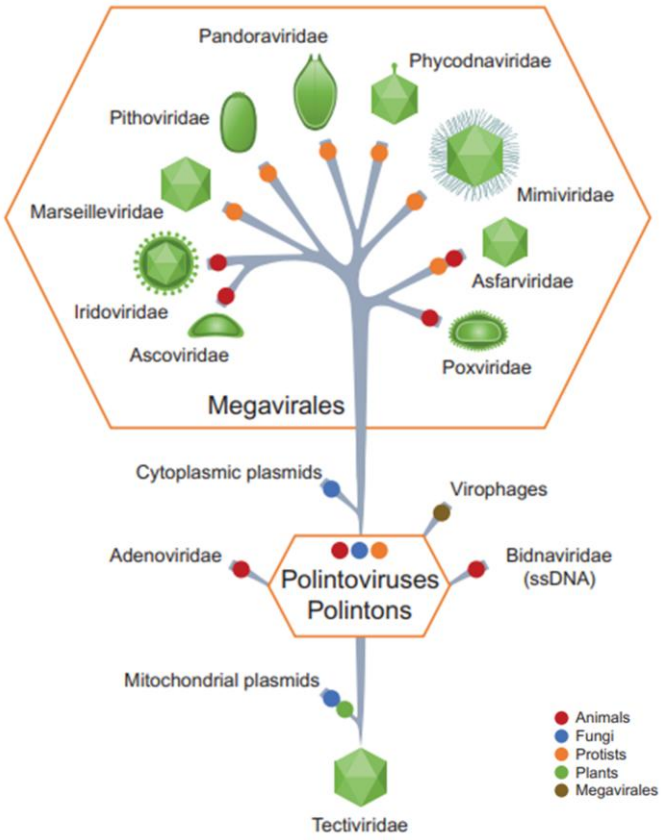

Figura 1. Esquema representativo da origem dos NCLDVs. Os políntons são prováveis intermediários entre ancestrais procarióticos, semelhantes ao Tectivírus, e os vírus de DNA dupla-fita, originando uma grande variedade de vírus com tamanhos amplamente diferentes, dentre eles os virófagos e os NCLDVs. **Fonte:** Koonin e Yutin, 2019 (Adaptado).

1.2 - A descoberta do primeiro vírus gigante de ameba e suas principais características

Em 1992, ocorreu um surto de pneumonia em um hospital da cidade de Bradford, na Inglaterra, no qual o agente etiológico era desconhecido. A fim de investigar o caso, foram realizadas coletas de amostras das torres de resfriamento de ar condicionado desse hospital e essas amostras foram analisadas pelo Dr. Tim Rowbothan e seus colaboradores no Laboratório de Saúde Pública em Leeds, Inglaterra. Rowbothan realizou a inoculação dessas amostras em culturas de *Acanthamoeba* sp. e ao observar o resultado do teste de Gram, comumente feito para identificação de bactérias, ele esperava encontrar bactérias gram-negativas que são comuns para pneumonia, como *Legionella* sp., entretanto, ele encontrou na verdade organismos que se

assemelhavam a cocos gram-positivos que foram chamados de “cocos de Bradford” por causa da cidade na qual foram isolados (Figura 2) (16).

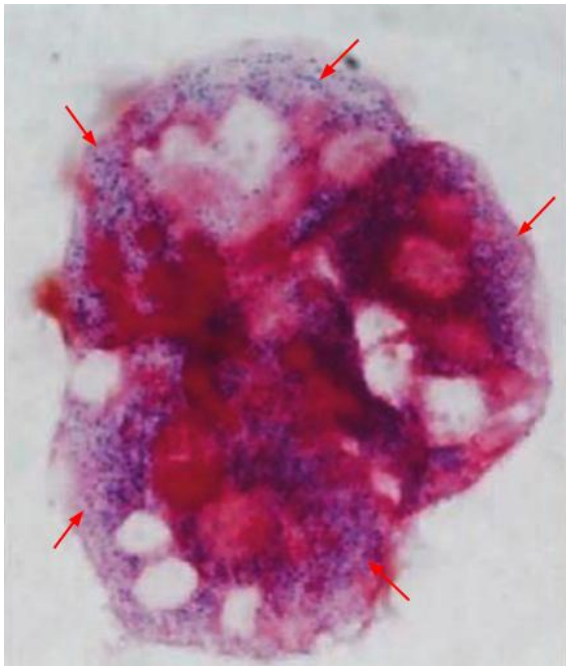

Figura 2. Visualização dos “cocos de Bradford”. Os chamados cocos de Bradford são esses pequenos pontinhos roxos (indicados pelas setas vermelhas), observados no interior de uma ameba após a coloração de Gram. **Fonte:** Raoult, La Scola e Birtles, 2007.

Inúmeras tentativas de identificação deste agente foram realizadas como: cultura e isolamento, testes com antibióticos, PCR e sequenciamento do gene 16S para bactérias, porém, todas as tentativas foram frustradas. Esse microrganismo foi então estocado e ficou esquecido por cerca de 10 anos, até que em 2003 um grupo de pesquisadores da Faculdade de Medicina da Aix Marseille Université da França, realizou uma microscopia eletrônica dessas amostras e se surpreendeu com o resultado. Eles descobriram que aquele agente era na verdade um vírus, não uma bactéria (Figura 3A). Esse resultado foi confirmado pela observação do ciclo de multiplicação desse organismo, que possui uma característica fase de eclipse. A análise do genoma completo confirmou a natureza viral desse microrganismo, que passou a ser chamado de Acanthamoeba polyphaga mimivírus (APMV). Esse nome foi escolhido devido ao isolamento ter sido realizado em amebas da espécie *Acanthamoeba polyphaga* e devido a semelhança desses vírus com bactérias gram-positivas na coloração de Gram, como se fosse uma forma de mimetismo (mimicking microbe), que fez com que os pesquisadores

confundissem os dois tipos de microrganismos (16).

O APMV foi o primeiro vírus gigante de amebas descoberto e ele faz parte atualmente da família *Mimiviridae*, gênero *Mimivirus*, espécie *Mimivirus bradfordmassiliense* (6). Seu capsídeo é formado por três camadas proteicas que revestem uma membrana interna lipídica, a qual circunda a parede do cerne (Figura 3A-B). O APMV apresenta uma simetria pseudo-icosaédrico com cerca de 450 nm de diâmetro e é circundado por fibrilas com cerca de 125 nm de diâmetro, formando uma partícula com 750 nm no total, o maior vírus identificado até aquele momento (Figura 3A; 3D). Além disso, esses vírus possuem uma região em forma de estrela em um dos vértices, denominada "stargate", que é por onde o genoma e proteínas de fase precoce são liberados durante a fase de desnudamento (Figura 3C) (5,17,18,19).

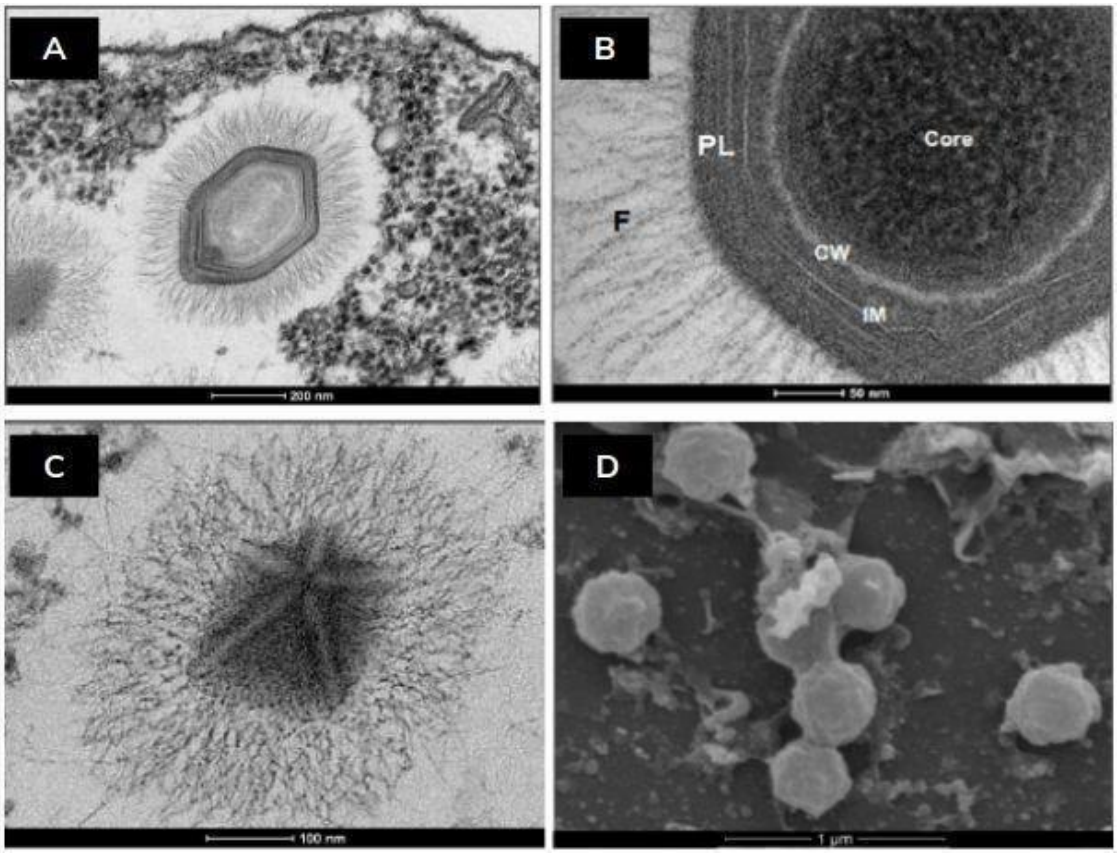

Figura 3. Imagens de um isolado de mimivírus. As partículas são observadas por Microscopia Eletrônica de Transmissão (MET) **(A, B e C)** e por Microscopia Eletrônica de Varredura (MEV) **(D)**. **(A)** Observação de um capsídeo pseudo-icosaédrico, circundado por fibrilas, com três camadas proteicas e uma membrana lipídica revestindo o cerne. **Barra**: 200 nm. **(B)** Maiores detalhes das estruturas que compõem a partícula dos mimivírus. **F:** Fibrilas; **PL:** Camadas proteicas; **IM:** Membrana interna; **CW:** parede do cerne. **Barra**: 50 nm. **(C)** Observação do vérice com o "stargate", estrutura pelo qual o genoma é liberado. **Barra:** 100 nm. **(D)** Observação da morfologia externa de partículas de um isolado de mimivírus através de um microscópio eletrônico de varredura. **Barra:** 1 µm. **Fonte:** Abrahão et al, 2014 (Adaptado) e Banco de imagens do GEPVIG.

As fibrilas que circundam o capsídeo são resistentes a protease, porém, quando tratadas com lisozimas elas se tornam sensíveis a estas enzimas, sugerindo assim, que elas estão imersas em uma matriz de uma substância similar a peptideoglicano, o que explica, em parte, o fato de reterem o cristal violeta durante a colocação de Gram. Estudos já demonstraram que essas fibrilas possuem um papel importante na adesão celular, apesar de sua ausência não impedir a multiplicação viral (20). Em 2011, Boyer e colaboradores realizaram 150 passagens do APMV em cultura axênica de *Acanthamoeba castellanii* resultando em uma variante denominada M4 que apresentou uma drástica redução das fibrilas, porém, o vírus mesmo assim é capaz de penetrar e se multiplicar na ameba hospedeira (21).

Como já foi citado, uma das características que chamou atenção dos pesquisadores com a descoberta do APMV foi a descrição de seu grande genoma. O genoma do APMV é composto por uma molécula de DNA linear fita dupla, com aproximadamente 1,2 megabases (Mb) que possui apenas 9,5% de DNA não codificante e mais de 1200 ORFs (*Open Reading Frames*). A partir das análises realizadas com o genoma do APMV, foram encontrados genes que nunca haviam sido descritos anteriormente para outros vírus como, genes que codificam enzimas envolvidas na tradução proteica (aminoacil-tRNA sintetase), proteínas envolvidas no reparo de DNA, outras proteínas como topoisomerases I e II, enzimas envolvidas na síntese de polissacarídeos, entre outras. Além disso, essas análises contribuíram para confirmar o agrupamento do mimivírus aos outros NCLDVs a partir da observação de diversos genes ortólogos entre esses vírus, incluindo os “core genes” (5).

1.3- Pimascovírus: Descoberta, principais características e contribuições para a expansão do conhecimento sobre a virosfera

1.3.1- Cedratvírus

O primeiro cedratvírus descrito, denominado Cedratvírus A11, foi isolado em 2016 em *Acanthamoeba castellanii* a partir de amostras coletadas em quatro regiões da Argélia. Esse vírus possui de 1-1,2 µm de comprimento e 0,5 µm de largura. A partícula apresenta um formato oval, seu capsídeo é estriado e possui dois poros apicais (Figura 4). O genoma do cedratvírus A11 é composto por uma molécula de DNA fita dupla circular com

aproximadamente 590.000 pares de bases e conteúdo GC de 42,6%. Foram preditos 574 genes, sendo que 177 destes (30,8%) foram definidos como ORFans (22).

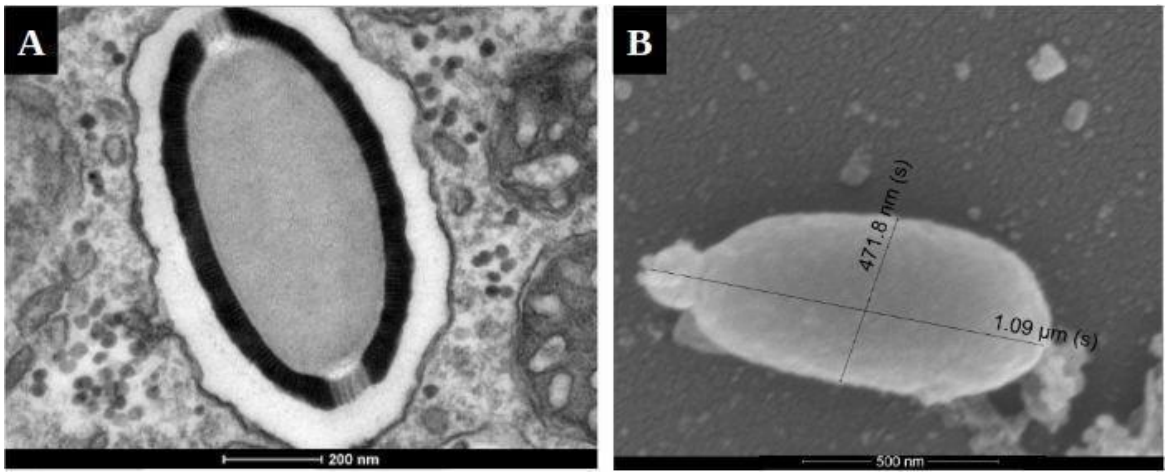

Figura 4. Imagens de um representante dos cedratvírus. (A) Morfologia interna de uma partícula de cedratvírus observada por microscopia eletrônica de transmissão. É possível observar seu capsídeo estriado e a presença dos dois poros apicais. **Barra**: 200 nm. **(B)** Morfologia externa da partícula observada por microscopia eletrônica de varredura. **Barra**: 500 nm. **Fonte:** Banco de imagens do GEPVIG.

Desde a descrição do cedratvírus A11, outros cedratvírus foram sendo isolados e caracterizados ao redor do mundo. Em 2017, foi descrito o isolamento do primeiro cedratvírus no Brasil. Este foi denominado Cedratvírus getuliensis, isolado em *A. castellanii* a partir de amostras de esgoto doméstico coletadas na cidade de Itaúna, Minas Gerais. Apesar de ser morfologicamente semelhante aos demais cedratvírus já isolados, a caracterização da partícula do cedratvírus getuliensis elucidou mais detalhes sobre o capsídeo viral. Além do capsídeo ser tipicamente estriado, foi observada a presença de uma membrana que delimita um compartimento interno (Figura 5) (23).

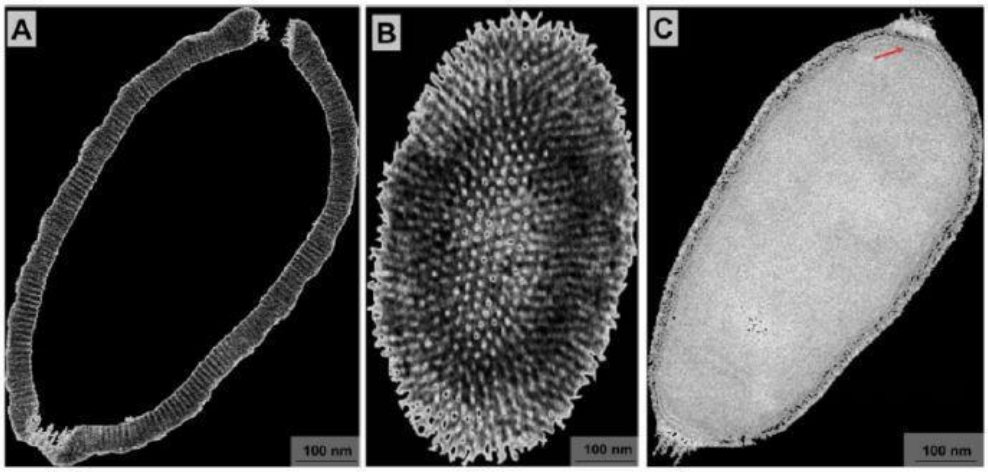

Figura 5. Características do capsídeo dos cedratvírus. (A) Capsídeo estriado com dois poros apicais. **Barra**: 100 nm **(B)** Listras paralelas do capsídeo se apresentam como pontos de forma organizada. **Barra**: 100 nm **(C)** O capsídeo apresenta uma membrana (seta vermelha) que delimita um compartimento interno. **Barra**: 100 nm. **Fonte:** Silva et al, 2018.

Em 2018, um novo isolado brasileiro denominado Brazilian cedratvírus trouxe novidade para o grupo dos cedratvírus. O brazilian cedratvírus possui uma partícula menor que os demais representantes do grupo, com aproximadamente 0,9 µm de comprimento e 0,5 µm de largura. Seu genoma também é o menor de todos os representantes dos cedratvírus descritos até o momento, com 460.000 pares de bases e codificando um número igualmente menor de proteínas, sendo 533 no total (24).

O isolamento do brazilian cedratvírus sugeriu a origem de uma nova linhagem (linhagem B) dentro do grupo dos cedratvírus. Foi realizada a construção de uma árvore filogenética baseada em sequências de aminoácidos do gene que codifica para a DNA polimerase B dos NCLDVs evidenciando a existência de duas linhagens diferentes dentro do grupo dos cedratvírus (Figura 6) (24).

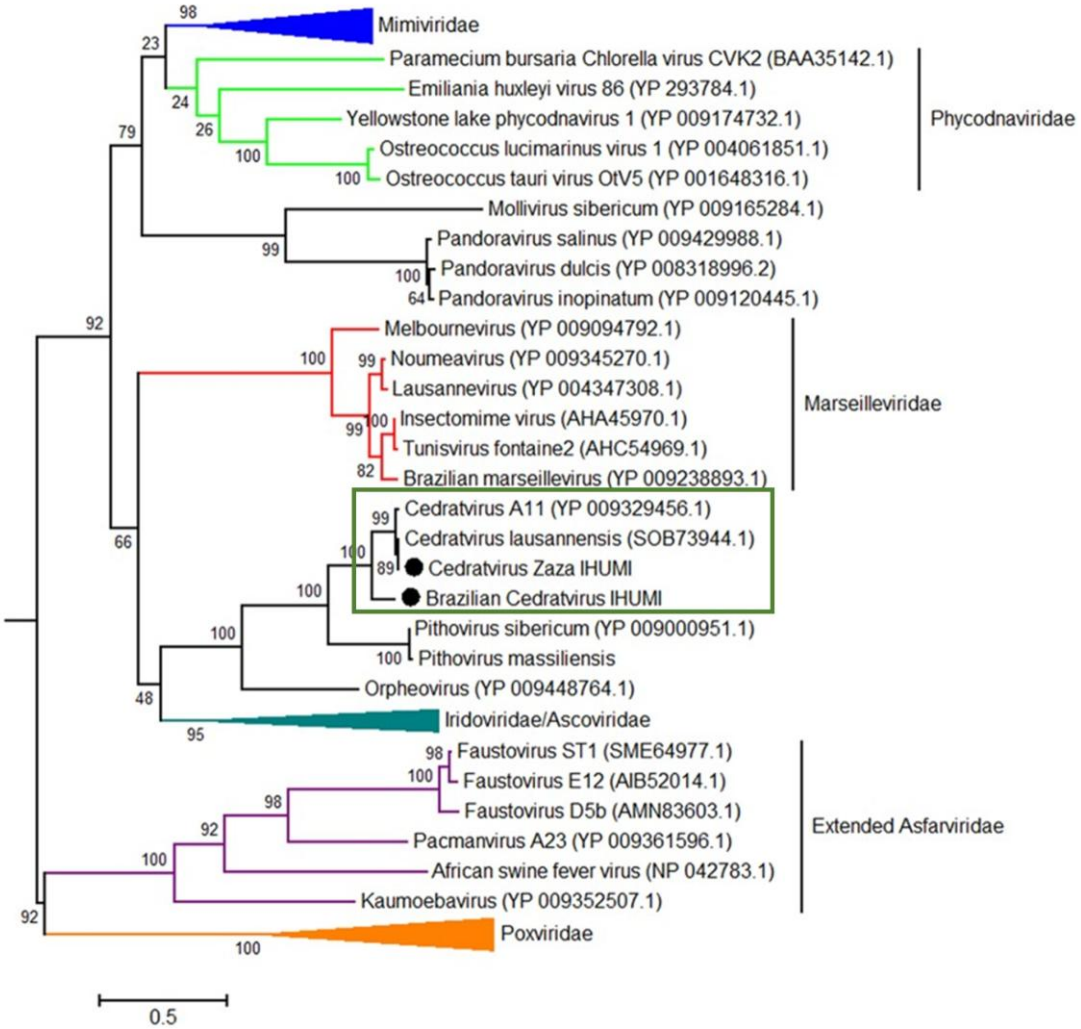

Figura 6. Árvore filogenética dos NCLDVs. Construída utilizando sequências de aminoácidos do gene que codifica para a DNA polimerase B. É possível observar a presença de uma segunda linhagem dentro do grupo dos cedratvírus (destacado pelo retângulo), separando Brazilian cedratvírus dos demais. **Fonte:** Rodrigues et al, 2018 (Adaptado).

Ciclo de multiplicação

A penetração das partículas de cedratvírus ocorre via fagocitose, 2 horas pós-infecção se inicia a fase de eclipse com a liberação do DNA viral. Posteriormente, as fábricas virais começam a ser formadas no citoplasma da célula hospedeira. As fábricas virais dos cedratvírus são elétron-lucentes, portanto, não é possível visualizar uma delimitação muito clara (Figura 7). São nessas fábricas virais que ocorrem os processos de replicação do DNA, morfogênese e montagem dos novos vírions. Após a morfogênese as partículas sofrem um espessamento do capsídeo ao passarem por uma região localizada na periferia da fábrica viral (Figura 8).

Durante a morfogênese é possível observar estruturas amorfas elétron-densas no citoplasma de algumas células (Figura 9). Essas estruturas também foram descritas por Legendre e colaboradores durante o ciclo de multiplicação do pithovírus sibericum (25) mas sua função ainda não é conhecida. Aproximadamente 12 horas pós-infecção as partículas são liberadas por lise celular. Antes da lise, algumas partículas provavelmente são liberadas por exocitose, porém, a via exocítica ainda precisa ser melhor estudada (22,23,24).

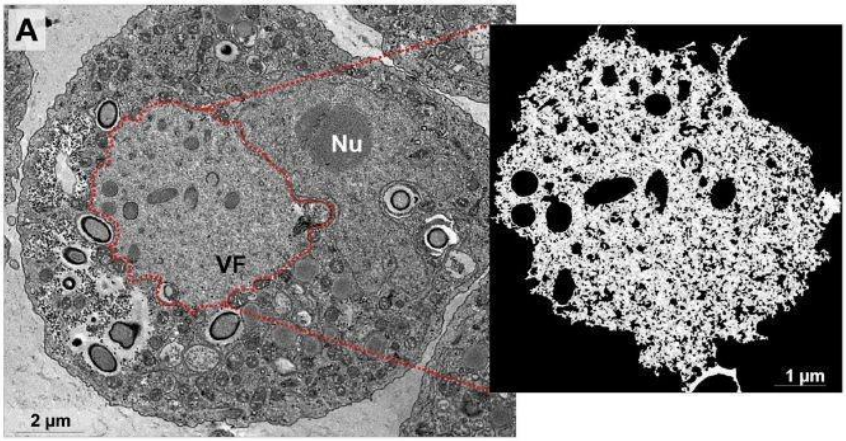

Figura 7. Fábrica viral do cedratvírus getuliensis. Imagem obtida por microscopia eletrônica de transmissão. Fábrica viral delimitada em vermelho e em destaque. **VF:** Fábrica viral. **Nu:** Núcleo. **Fonte:** Silva et al, 2018.

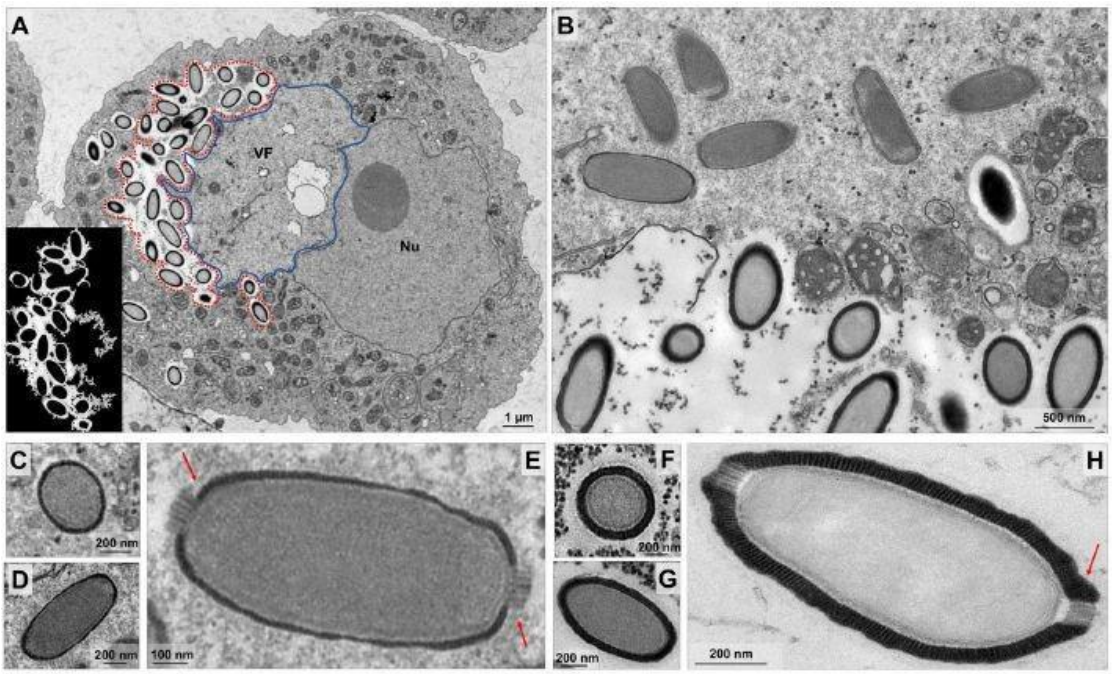

Figura 8. Imagens de MET demonstrando o espessamento das partículas do cedratvírus getuliensis. (A) Partículas virais sofrem espessamento do capsídeo em uma região (contornada em vermelho) na periferia na

fábrica viral (contornada em azul), **Barra**: 1 µm. **(B)** Periferia da fábrica viral com partículas já espessadas (abaixo) e partículas que ainda não passaram pelo processo de espessamento (acima), **Barra:** 500 nm. Secções transversais **(C)** e longitudinais **(D)** evidenciando um capsídeo com parede fina que ainda não recobre os poros (setas vermelhas) **(E)**. Secções transversais **(F)** e longitudinais **(G)** evidenciando o espessamento do capsídeo, que agora recobre os poros (setas vermelhas) **(H)**. **VF**: Fábrica viral. **Nu**: Núcleo. **Fonte:** Silva et al, 2018.

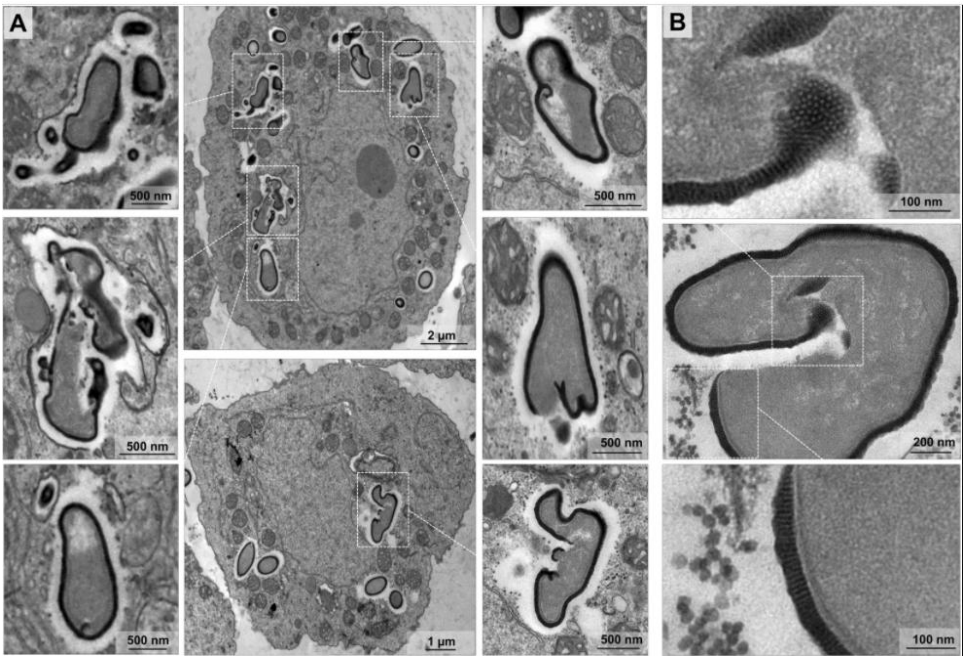

Figura 9. Estruturas amorfas observadas durante o ciclo de multiplicação do cedratvírus getuliensis. (A) e **(B)** Estruturas amorfas de material elétron-denso compostas por porções de corks, capsídeo estriado e, membrana visualizadas no citoplasma de algumas células hospedeiras. **Fonte:** Silva et al, 2018.

1.3.2- Pithovírus

Em 2014, Legendre e colaboradores isolaram um novo grupo de vírus gigantes em *Acanthamoeba* que foi denominado pithovírus. O pithovírus sibericum foi isolado na Sibéria a partir de amostras de *permafrost* datado de 30.000 anos (25).

Assim como os cedratvírus, os pithovirus possuem partículas ovais. Porém, suas partículas são maiores, com cerca de 1,3 µm em média, e seu capsídeo é composto por estrias verticais paralelas, incluindo o poro apical (ou cork) (Figura 10). Se comparado ao grande tamanho de sua partícula, o genoma é relativamente pequeno, composto por DNA fita dupla circular com aproximadamente 610 kpb codificando cerca de 467 proteínas (25).

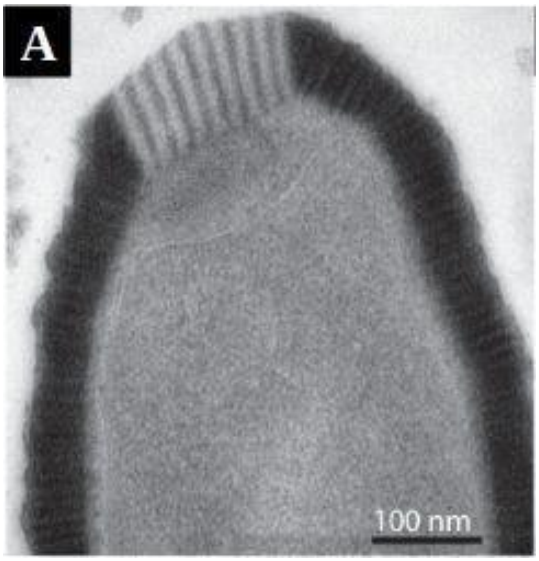

Figura 10. Imagem de uma partícula de pithovírus. Imagem obtida por microscopia eletrônica de transmissão. O capsídeo e o poro apical são compostos por estrias paralelas. **Barra:** 100 nm. **Fonte:** Legendre et al, 2014.

1.3.3- Orpheovírus

Um novo vírus isolado em 2018 foi agrupado com os cedratvírus e pithovírus, esse novo isolado foi nomeado de Orpheovírus IHUMI. Esse vírus foi isolado em *Vermamoeba vermiformis* a partir de uma amostra de fezes de rato coletadas em La Ciotat, na França. Esses vírus apresentam partículas ovais de 900-1300 nm e possuem um poro apical, semelhante ao dos pandoravírus (Figura 11, Figura 12F). Seu genoma possui aproximadamente 1,47 Mb codificando cerca de 1512 proteínas (26).

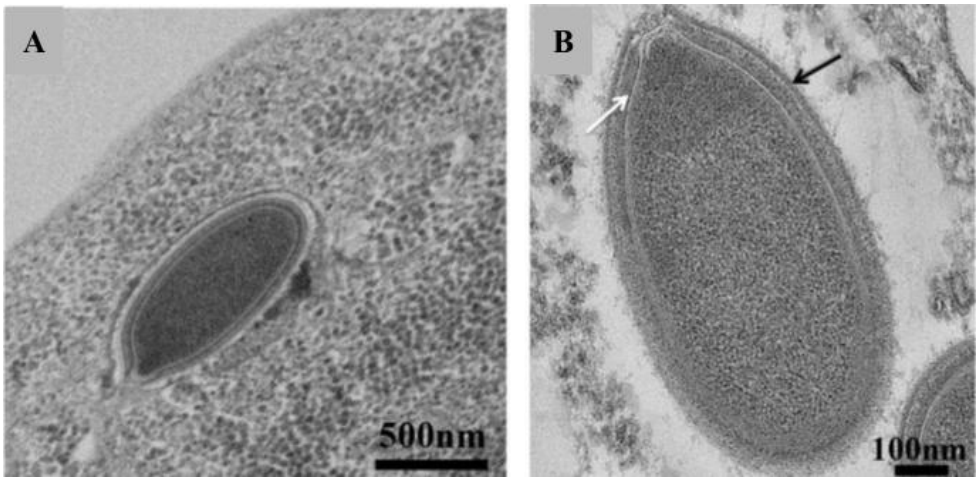

Figura 11. Partículas de Orpheovírus. (A) Observação por MET de uma partícula de Orpheovírus no interior da célula hospedeira. **Barra**: 500 nm. **(B)** Partícula madura de Orpheovírus, evidenciando a membrana externa (seta preta) e o espaço denso médio (seta branca). **Barra:** 100 nm. **Fonte:** Andreani et al, 2018, adaptado.

1.4- Isolamento e distribuição de outros vírus gigantes no mundo

Além das famílias e dos grupos virais já citados, vários outros grupos de vírus gigantes de amebas foram isolados ao longo dos últimos 20 anos (Figura 12).

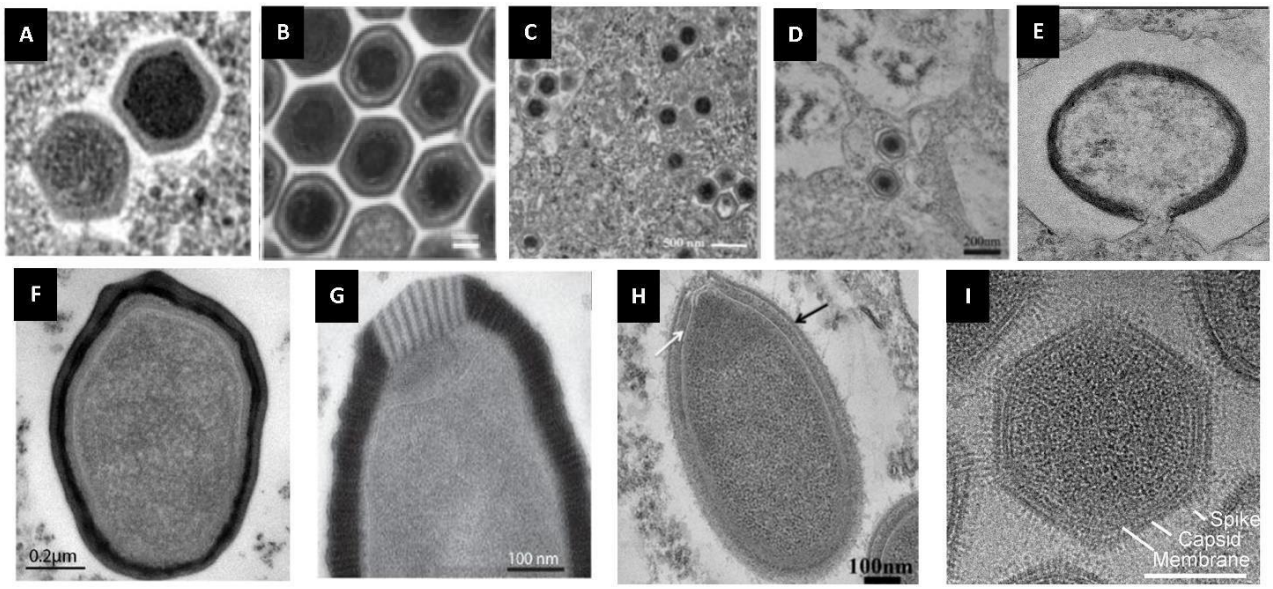

Figura 12. Os diferentes grupos de vírus gigantes. Grupos de vírus gigantes isolados após a descoberta dos mimivírus. **(A)** Marseillevírus. **(B)** Faustovírus. **Barra:** 100 nm. **Barra:** 100 nm. **(C)** Kaumoebavírus. **Barra:** 500 nm. **(D)** Pacmanvírus. **Barra:** 200 nm. **(E)** Mollivírus. **(F)** Pandoravírus. **Barra:** 200 nm. **(G)** Pithovírus. **Barra:** 100 nm. **(H)** Orpheovírus. A seta preta indica a membrana externa e a seta branca indica o espaço denso médio. **Barra:** 100 nm. **(I)** Medusavírus. Estão indicadas as posições da membrana, do capsídeo e da estrutura spike. **Barra:** 100 nm. **Fonte:** Boyer et al, 2009; Reteno et al, 2015; Bajrai et al, 2016; Andreani et al, 2017; Legendre et al, 2015; Philippe et al, 2013; Legendre et al, 2014; Andreani et al, 2018; Yoshikawa et al, 2019.

Os vírus gigantes são muito diversos em diferentes aspectos. A morfologia desses vírus pode ser icosaédrica (ex. marseillevírus), pseudoicosaédrica (mimivírus), oval (ex. pandoravírus), esférica (mollivírus) ou pseudoicosaédrica com cauda (tupanvírus) e seus capsídeos podem ser recobertos por fibrilas ou até mesmo por espículas (medusavírus). Além de características morfológicas, esses vírus também apresentam diferenças em relação ao

genoma, seja devido ao seu tamanho ou sua composição. Os vírus gigantes são ubíquos, tendo sido isolados de diferentes tipos de amostras em vários lugares do mundo. Na Tabela 1 são apresentadas algumas informações sobre os primeiros isolados de cada grupo conhecido até o momento para os vírus gigantes de amebas. Vale ressaltar que, para alguns desses grupos, muitos isolados já foram descritos e, portanto, a diversidade é ainda maior. Entretanto, já é possível observar um pouco dessa diversidade encontrada para esses vírus, mesmo sendo descritos apenas os primeiros isolados.

Tabela 1. Grupos de vírus gigantes isolados em amebas e algumas características principais. A tabela apresenta os grupos de vírus gigantes conhecidos até o momento e informações resumidas relacionadas ao primeiro isolado de cada grupo.

Vírus	Hospedeiro	Tamanho do capsídeo	Tamanho do genoma (Kb)	Amostras de isolamento	País de isolamento	Referência
Mimivírus	*Acanthamoeba sp.*	750 nm	1.200	Água de torre de resfriamento	Inglaterra	(4,5)
Marseillevírus	*Acanthamoeba sp*.	250 nm	370	Água de torre de resfriamento	França	(27)
Pandoravírus	*Acanthamoeba sp.*	1 µm	1.900-2.800	Sedimento marinho e lama de água doce	Chile e Austrália	(28)
Pithovírus	*Acanthamoeba sp.*	1,5 µm	610	Permafrost	Rússia	(25)
Faustovírus	*Vermamoeba vermiformis*	200 nm	466	Esgoto e água do mar/sedimento	França e Senegal	(29)
Mollivírus	*Acanthamoeba sp.*	500-600 nm	651	Permafrost	Rússia	(30)
Cedratvírus	*Acanthamoeba sp.*	1-1,2 µm	590	Amostra ambiental	Argélia	(22)
Kaumoebavírus	*Vermamoeba vermiformis*	250 nm	350	Esgoto	Arábia Saudita	(31)
Pacmanvírus	*Acanthamoeba sp.*	250 nm	395	Amostra ambiental	Argélia	(32)
Orpheovírus	*Vermamoeba vermiformis*	900-1300 nm	1473	Fezes de rato	França	(26)
Medusavírus	*Acanthamoeba sp.*	250 nm	381	Água termal	Japão	(33)

Todos esses achados demonstram o quanto esses vírus são diversos e estão presentes no mundo inteiro, além disso reforçam a importância dos estudos de prospecção para o enriquecimento do conhecimento sobre esses vírus.

1.5- O gigantismo genômico dos vírus gigantes

Conforme supracitado, a descrição do APMV chamou atenção não apenas devido a sua grande partícula, mas também devido ao seu grande genoma (4,5,34). Com o isolamento de novos vírus gigantes, nos anos posteriores, mais vírus gigantes com grandes genomas foram sendo descobertos. Embora os mecanismos que influenciaram no gigantismo do genoma desses vírus sejam desconhecidos, a transferência horizontal de genes, a emergência de genes de novo e a duplicação de genes têm sido apontadas como possíveis hipóteses (27,34,35).

A transferência gênica horizontal é um processo biológico pelo qual genes ou fragmentos de material genético são transferidos de um organismo para outro. Por ser um processo independente de reprodução sexuada, pode ocorrer entre organismos de espécies diferentes. A TGH é muito comum entre bactérias e é fundamental para sua evolução e diversificação, sendo responsável, por exemplo, pela transferência de genes de resistência a antibióticos e genes relacionados à virulência (36).

Dentre os vírus gigantes, a TGH já foi descrita para os marseillevírus que apresentam um grande mosaicismo genômico, uma vez que esse grupo de vírus possui genes que codificam para proteínas que se mostraram altamente similares a sequências homólogas de outros vírus, bactérias, eucariotos e arqueias (27,37). Esse mosaicismo pode ser explicado devido ao estilo de vida simpátrico das amebas, que são os prováveis hospedeiros desses vírus, uma vez que elas são de vida livre e capazes de fagocitar diferentes organismos, possibilitando a transferência de genes entre os diferentes grupos de organismo durante o processo de replicação do DNA viral.

Outro mecanismo de expansão genômica é a geração de genes de novo. A origem de genes de novo, refere-se ao processo pelo qual novos genes funcionais surgem a partir de sequências de DNA não codificantes ou de origem não gênica. Esse mecanismo já foi sugerido para os pandoravírus, e parece ser um mecanismo importante para a expansão do genoma desse vírus (35).

Finalmente, o terceiro mecanismo proposto relacionado à expansão do genoma em vírus gigantes é a duplicação de genes. A duplicação de genes é um mecanismo crítico subjacente à evolução dos organismos. Sua importância reside na geração de matéria-prima para a evolução de novas funções e vias gênicas (38). Genes duplicados podem adquirir mutações, resultando na evolução de novas funções proteicas. A importância desse fenômeno na evolução é evidenciada pela ocorrência generalizada de genes duplicados em todos os domínios da vida

(39,40,41,42). A duplicação tem sido o mecanismo responsável pela produção de cerca de 30-65% dos genes em eucariotos multicelulares, como os humanos (43,44).

Embora os estudos sobre duplicação de genes em vírus gigantes sejam escassos, há evidências convincentes de que aproximadamente um terço do genoma do mimivírus e 50% dos genomas dos Pandoravírus são compostos por genes em cópias múltiplas (35,34).

1.6 – O estudo da proteômica em vírus gigantes

Os vírus gigantes surpreenderam a comunidade científica por seu tamanho excepcional e genomas complexos, desafiando a fronteira tradicional entre vírus e organismos celulares.

Análises genômicas de diversos vírus gigantes revelam a presença de uma alta prevalência de ORFans (*Open Reading Frames*). São chamadas de ORFans os genes que codificam proteínas com funções desconhecidas e que não são encontradas em nenhum banco de dados conhecido. Por exemplo, no genoma do Acanthamoeba polyphaga mimivirus (APMV), as ORFans representam aproximadamente 67% do genoma viral (5), enquanto no Pandoravirus salinus, eles constituem cerca de 93% (28).

Apesar de escassos, os estudos de proteômica de vírus gigantes revelam a presença de uma gama de proteínas que são essenciais em diversos processos celulares. As proteínas relacionadas à replicação e reparo de DNA, transcrição e processamento de RNA, por exemplo, sugerem que esses vírus são capazes de iniciar a replicação e transcrição de seu genoma logo após infectar seu hospedeiro de forma mais autônoma (19,25,27,28,45,46,47). Foram identificadas também topoisomerases, que são enzimas responsáveis por regular o superenrolamento do DNA e podem atuar na transcrição de genes precoces, logo após o desnudamento viral (45). Além disso, várias proteínas associadas ao metabolismo e enzimas redox foram identificadas (25,28,45), sugerindo a potencial presença de mecanismos para lidar com o estresse oxidativo do hospedeiro e manipulação do ambiente hospedeiro no início do ciclo.

Os estudos de proteômica para vírus gigantes, apesar de limitados, são essenciais para expandir nossa compreensão acerca dos papéis ecológicos que esses vírus desempenham na natureza, além de mapear suas interações com o hospedeiro, contribuindo para o avanço na compreensão da evolução viral. Esses estudos são importantes também para confirmar a

existência de ORFans previstos pela bioinformática, ajudando a estabelecer sua presença na partícula viral e orientando futuras investigações funcionais.

2 - JUSTIFICATIVA

Após o isolamento do *Acanthamoeba polyphaga mimivirus* (APMV) em 2003, primeiro membro da família *Mimiviridae*, a busca por novos vírus gigantes que infectam amebas foi intensificada. Essa busca resultou na descoberta de vários outros grupos de vírus gigantes, isolados dos mais diversos ambientes ao redor do mundo.

Ao serem descobertos, esses vírus chamaram muita atenção devido ao grande tamanho de suas partículas, mas também devido ao seu grande e complexo genoma. Alguns vírus gigantes possuem genomas maiores até mesmo do que de algumas bactérias. Os fenômenos responsáveis pela expansão do genoma destes vírus ainda não foram elucidados de forma aprofundada, entretanto, três mecanismos têm sido propostos e parecem ser relevantes dependendo do vírus em questão: transferência gênica horizontal, criação de genes *de novo* e duplicação de genes. Estudar mais profundamente e entender esses mecanismos nos ajuda a entender a história evolutiva em cada um dos grupos.

Apesar de escassos, os estudos de proteômica de vírus gigantes publicados até o momento, identificam a presença de proteínas chave que são essenciais para diversas atividades celulares. A presença de proteínas relacionadas à replicação e reparo de DNA, transcrição e processamento de RNA no capsídeo, por exemplo, indicam um certo nível de autonomia viral, e nos ajuda a entender aspectos básicos do ciclo deste vírus, como por exemplo seus sítios celulares de propagação. Neste contexto, o presente trabalho realizou análises biológicas e genômicas de um novo cedratvírus isolado previamente, o cedratvírus pambiensis. Nossas análises genômicas revelaram uma grande quantidade de eventos de duplicação de genes, e análises mais aprofundadas nos forneceram informações sobre como essas duplicações acontecem, como estão distribuídas ao longo do genoma e, ainda, se estão presentes também em outros nucleocitovírus. Todos esses resultados nos ajudam a elucidar a influência desse mecanismo no gigantismo genômico de alguns vírus gigantes e como acontece a evolução dos genomas desses vírus.

Neste trabalho também foi realizada uma análise proteômica de partículas purificadas do cedratvirus pambiensis. Entender a diversidade de proteínas presentes no capsídeo de um vírus também é muito importante para entender aspectos relacionados aos mecanismos de penetração e manipulação da maquinaria celular do hospedeiro, por exemplo.

3 - OBJETIVOS

3.1- Objetivo geral

Caracterizar o genoma, a evolução, o ciclo e as partículas de um novo vírus gigante isolado no Brasil, denominado cedratvirus pambiensis

3.2- Objetivos específicos

- Caracterizar morfologicamente o cedratvírus pambiensis, utilizando microscopia eletrônica de transmissão, microscopia eletrônica de varredura e microscopia eletrônica por contrastação negativa;
- Analisar as etapas do ciclo de multiplicação do cdv. pambiensis utilizando ensaios de ciclo único e microscopia eletrônica de transmissão;
- Sequenciar e montar o genoma do cdv. pambiensis;
- Realizar análises filogenéticas;
- Realizar a predição de genes; anotar e caracterizar funcionalmente as proteínas preditas;
- Analisar características evolutivas e estruturais do genoma do cdv. pambiensis e expandir as análises para outros nucleocitovírus;
- Identificar e analisar as proteínas presentes nas partículas purificadas do cdv. pambiensis;
- Comparar os resultados obtidos do proteoma de cdv. pambiensis com o proteoma publicado de pithovirus sibericum.

4 – METODOLOGIA, RESULTADOS E DISCUSSÕES PARCIAIS SERÃO APRESENTADOS A SEGUIR NA FORMA DE ARTIGOS PUBLICADOS PRECEDIDOS DE UM BREVE RESUMO.

4.1 – Artigo 1 – Isolation of Giant viruses of *Acanthamoeba castellanii*

Este artigo foi publicado no periódico Current Protocols

Este artigo descreve um método prático para prospecção e isolamento de vírus gigantes com base na inoculação direta de amostras ambientais em culturas de amebas de *Acanthamoeba castellanii*. Os vírus gigantes que infectam amebas já foram isolados de várias amostras ambientais em vários países do mundo, incluindo em ambientes extremos. Aqui descrevemos os procedimentos metodológicos relativos à prospecção de vírus gigantes em *A. castellanii*, incluindo a preparação de amostras ambientais, a cultura de amebas e a observação de efeitos citopáticos que podem indicar a presença e o potencial isolamento de vírus gigantes.

Isolation of Giant Viruses of *Acanthamoeba castellanii*

Talita Bastos Machado,[1] Isabella Luiza Martins de Aquino,[1] and Jônatas Santos Abrahão[1,2]

[1]Instituto de Ciências Biológicas, Departamento de Microbiologia, Laboratório de Vírus, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[2]Corresponding author: *jonatas.abrahao@gmail.com*

This article describes a practical method for prospecting and isolating giant viruses based on direct inoculation of environmental samples into amoeba cultures of *Acanthamoeba castellanii*. The giant viruses that infect amoebas have already been isolated from various environmental samples in several countries worldwide, including in extreme environments. Here we describe the methodologic procedures regarding the prospecting of giant viruses in *A. castellanii*, including the preparation of environmental samples, the culture of amoebas, and the observation of cytopathic effects that can indicate the presence and potential isolation of giant viruses. © 2022 Wiley Periodicals LLC.

Basic Protocol 1: Sample collection
Support Protocol: Propagation of *Acanthamoeba castellanii*
Basic Protocol 2: Prospecting of giant viruses in environmental samples by cytopathic effect analysis

Keywords: *Acanthamoeba castellanii* • cytopathic effect • giant viruses • virus prospecting

How to cite this article:
Machado, T. B., de Aquino, I. L. M., & Abrahão, J. S. (2022). Isolation of giant viruses of *Acanthamoeba castellanii*. *Current Protocols*, *2*, e455. doi: 10.1002/cpz1.455

INTRODUCTION

Since the discovery of *Acanthamoeba polyphaga* mimivirus (APMV) in 2003 (La Scola et al., 2003), several other groups of giant viruses that infect amoebas have been isolated: marseillevirus, pandoravirus, pithovirus, faustovirus, mollivirus, cedratvirus, kaumoebavirus, pacmanvirus, orpheovirus, and medusavirus (Andreani et al., 2016, 2017, 2018; Bajrai et al., 2016; Boyer et al., 2009; Legendre et al., 2014, 2015; Philippe et al., 2013; Reteno et al., 2015; Yoshikawa et al., 2019). These findings have contributed to the expansion of our knowledge regarding the so-called nucleocytoplasmic large DNA viruses. This group includes several virus families (e.g., *Poxviridae* and *Phycodnaviridae*) in addition to these giant amoebal viruses. The giant viruses of amoebas have already been isolated from different types of samples, such as soil, mud, sewage, spring water, seawater, ocean sediments, and even clinical samples. They have been isolated in several countries worldwide from different environments, including extreme ones such as the ocean floor, saline lagoons, and even Antarctica (Abrahão et al., 2018; Andrade et al., 2018). The abundant presence of giant virus sequences in metagenomics databases may indicate that these viral entities, as well as their hosts, are ubiquitous (Kerepesi & Grolmusz., 2016, 2017; Mihara et al., 2018; Schulz et al., 2017, 2018; Wu et al., 2020).

CURRENT PROTOCOLS
A Wiley Brand

Current Protocols e455, Volume 2
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpz1.455

Overall, giant viruses present a particularly narrow host range. Most of them were isolated from amoebas of the genus *Acanthamoeba*, mostly *A. castellanii*, including marseillevirus, pandoravirus, pithovirus, mollivirus, cedratvirus, pacmanvirus, and medusavirus. Giant viruses have also been isolated from the amoeba species *Vermamoeba vermiformis*, which is known to be permissive to the replication of faustovirus, kaumoebavirus, and orpheovirus. Although members of the family *Mimiviridae* also have a narrow host range, notable known exceptions are the tupanviruses, which are able to infect different genera of protists including *Acanthamoeba*, *Vermamoeba*, *Willartia*, and *Dictyostelium* (Abrahão et al., 2018). Nevertheless, *A. castellanii* has been the most popular platform to isolate giant viruses, since this organism can be easily propagated and is permissive to a great diversity of giant viruses.

Here we describe how to isolate giant viruses of *A. castellanii*. This species was chosen because it is the most generalist host for giant viruses among amoeba species. The protocol for maintaining these amoebas is described, as is the protocol for preparing 96-well microplates for viral prospecting. We also describe how each type of environmental sample must be prepared before processing and preventive measures to avoid contamination with other microorganisms such as fungi and bacteria.

BASIC PROTOCOL 1

SAMPLE COLLECTION

Samples can be collected from different environments, including soil, sewage, mud, and fresh and marine water.

Materials

Sample of interest
15- and 50-ml conical centrifuge tubes

NOTE: All materials must be sterile. Gloves must be used to collect the samples.

Water and sewage samples (without sediment)

1a. Immerse tubes in liquid until collecting ≥10 ml.

Better results have been obtained when water samples were collected at the surface of creeks and shallow polluted lagoons.

Mud samples (water with sediment)

1b. Collect sediment along with the liquid until ≥10 ml liquid is collected.

Soil samples (sediment only)

1c. Collect soft and humid soil pushing the borders of the sterile tube onto the surface of the ground.

For soil samples, 1 g is sufficient.

2. Transport samples at room temperature, and store at 4°C for up to 1 month until the next step.

SUPPORT PROTOCOL

PROPAGATION OF *Acanthamoeba castellanii*

In this protocol, amoebas of the species *A. castellanii* are presented and used as an isolation platform owing to its permissivity to a substantial number of known giant viruses.

Materials

A. castellanii cells (e.g., ATCC #30234)
Peptone-yeast extract with glucose (PYG) medium (see recipe)

25-, 75-, and 175-cm^2 tissue culture flasks
Inverted microscope

50-ml conical centrifuge tubes
Neubauer chamber and glass cover (for cell counting)
1.5-ml microcentrifuge tubes
30°C incubator

NOTE: This procedure must be performed in a biosafety level 2 (BSL-2) cabinet. All solutions and materials must be sterile.

1. Split fresh *A. castellanii* cells in 75-cm^2 tissue culture flasks containing 15 ml PYG medium at 30°C. First, under an inverted microscope, observe if the number of amoebas is adequate to undergo cell splitting (about 90% to 100% cell confluence).

   When the tissue culture flask is too full of amoebas, cells begin to detach from the monolayer, so it is best to split closer to 90% cell confluence. After 2 or 3 days, the split can be performed again, according to the number of cells inoculated previously.

2. Discard medium from the tissue culture flask, and add 10 ml fresh PYG medium.

   Cell splitting occurs only with amoebas that are attached to the culture flask.

3. To detach cells, gently hit the flask with the palm of the hand. Subsequently, gently homogenize cells with a pipette.

   Unlike other cells, amoebas detach from the monolayer mechanically; trypsin is not required. Therefore, the tissue culture flasks must be handled carefully.

4. Transfer cells to 50-ml conical centrifuge tubes, and perform cell counting using a Neubauer chamber. For counting, prepare a 1:10 dilution of cells in a 1.5-ml microcentrifuge tube. Homogenize solution thoroughly with a micropipette before counting. Load Neubauer chamber with 10 µl dilution using a micropipette.

5. Subculture cells in a new tissue culture flask adding 500,000; 1,000,000; and 2,500,000 cells to 25-, 75-, and 150-cm^2 tissue culture flasks containing a final volume of 5, 15, and 25 ml PYG medium, respectively.

6. Incubate at 30°C until cell confluence (about 90% to 100%) is reached.

   Perform cell splitting weekly, at least three times.

PROSPECTING OF GIANT VIRUSES IN ENVIRONMENTAL SAMPLES BY CYTOPATHIC EFFECT ANALYSIS

BASIC PROTOCOL 2

It is necessary to realize different preparations for each type of sample, according to its nature (water, mud, or soil). For all samples, up to 10 aliquots of 1 ml should be made. In order to prospect giant viruses in this protocol, 96-well microplates are used, which allows for the processing of several samples at the same time. This prospection technique consists of a coculture method using 100 µl *A. castellanii* (4×10^4 cells) and 100 µl of each environmental sample per well.

Materials

Sample of interest (see Basic Protocol 1)
Sterile water
PYG medium (see recipe)
Vancomycin
Ciprofloxacin
Doxycycline
75-cm^2 tissue culture flasks containing *A. castellanii* (see Support Protocol)
1× phosphate-buffered saline (PBS) (see recipe)

1.5-ml microcentrifuge tubes, sterile
Inverted microscope

50-ml conical centrifuge tubes
Neubauer chamber and glass cover (for cell counting)
96-well microplate
Tape (to seal microplate)
30°C incubator
25-cm^2 tissue culture flask

NOTE: This procedure must be performed in a BSL-2 cabinet. All solutions and materials must be sterile.

Sample preparation

Water and sewage samples (without sediment)

1a. Aliquot contents collected in conical centrifuge tubes into 1.5-ml microcentrifuge tubes (1 ml per tube), and perform a 1:10 dilution of each aliquot in sterile water.

Mud samples (water with sediment)

1b. Refrigerate collected material for 24 hr to allow any sediment to settle. Subsequently, prepare 1-ml aliquots with the supernatant in 1.5-ml microcentrifuge tubes. Finally, prepare a 1:10 dilution with each aliquot in sterile water.

Soil samples (sediment only)

1c. Dilute collected soil in sterile water (1:10) and homogenize thoroughly. Subsequently, store solution at 4°C for 24 hr to allow for sedimentation. Then, aliquot supernatant into 1.5-ml microcentrifuge tubes (1 ml per tube).

All aliquots must undergo three freeze-thaw cycles prior to processing (in order to avoid contamination by fungi and bacteria). Store aliquots in freezer boxes at –20°C.

96-well microplate preparation

2. Supplement PYG medium with additional antibiotics (0.004 mg/ml vancomycin, 0.004 mg/ml ciprofloxacin, and 0.02 mg/ml doxycycline) to reduce potential contamination from environmental samples to be prospected.
3. Observe culture flasks under an inverted microscope to check cell confluence (ideally about 90% to 100%).
4. Discard medium from the tissue culture flask, and add 10 ml fresh PYG medium supplemented with additional antibiotics.

The procedure is carried out only with amoebas that are attached to the culture flasks.

5. To detach cells, gently hit the flask with the palm of the hand (as described in Support Protocol step 3).
6. Transfer cells to 50-ml conical centrifuge tubes, and perform cell counting using a Neubauer chamber. For counting, prepare a 1:10 dilution of cells in 1.5-ml microcentrifuge tubes and homogenize thoroughly. Load Neubauer chamber with 10 µl dilution using a micropipette.
7. In 96-well microplates, add 40,000 cells per well with a final volume of 100 µl per well.
8. Seal edges of the microplates with masking tape, and incubate at 30°C until inoculation of samples.

Sample inoculation

A total of three passages is performed.

9. Passage cells using sample and appropriate control. For each sample, include wells with the original sample (not diluted) and another with the sample diluted. Use the

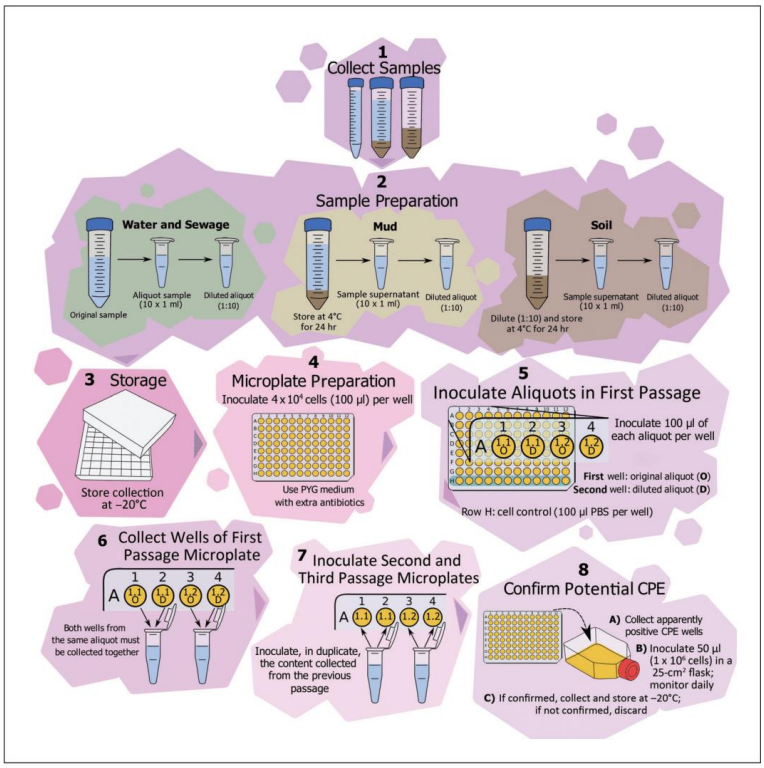

Figure 1 Complete pipeline of the prospecting processes. The process starts with the collection of samples (1). Different preparations are needed according to the nature of the sample (2). Collections are then organized in freezer boxes (3). The 96-well microplates must be prepared (4) to perform the three prospecting passages (5, 6, and 7). Finally, if CPE are observed in any well, they must be confirmed (8). CPE, cytopathic effects; PBS, phosphate-buffered saline; PGY, peptone-yeast extract with glucose.

entire row H (see Fig. 1, item 5) for a cell-only control, which consists of 100 µl *A. castellanii* (4×10^4 cells) and 100 µl PBS.

10. At the end of the first passage, collect original (not diluted) and diluted aliquots of each sample, and transfer to a single 1.5-ml microcentrifuge tube (see Fig. 1, item 6). Use this mixture in the next passage, and keep at –20°C until inoculation on a new microplate.

11. Perform the second and third passages with no further dilutions. Use the mixture of the two aliquots (from the same sample) collected from the previous passage in the next passage.

    For each passage, the microplate must be sealed with tape to prevent the lid from opening and incubated at 30°C for 7 days, being visually inspected daily for cytopathic effects (CPE).

12. If any of the aliquots among the three passages presents CPE, collect contents of the wells, and store at –20°C for further identification.

    This procedure must be performed in a BSL-2 cabinet, and all materials must be sterile.

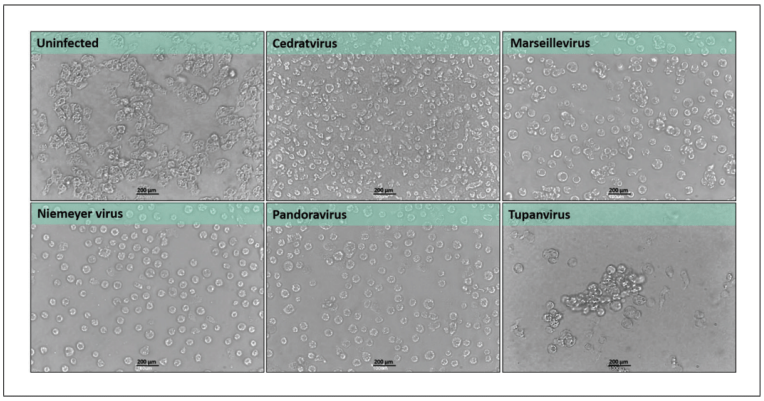

Figure 2 Cytopathic effects caused by different viral groups in *Acanthamoeba castellanii*. Most giant viruses cause cell rounding, detaching of the cell monolayer, and ultimately cell lysis. The formation of clusters or bunches is observed in cells infected by tupanviruses. Adapted and reprinted from de Souza et al. (2020) under Creative Commons Attribution 4.0 International License.

CPE may vary slightly between viral groups. However, the most common effect is cell rounding and detachment of the monolayer, followed by cell lysis. Tupanvirus induces a different CPE, which is the formation of clusters or bunches of rounded cells (Fig. 2). To date, this CPE has only been observed for the two tupanvirus isolates.

Confirmation of cytopathic effects

13. Add 1×10^6 *A. castellanii* cells to a 25-cm^2 tissue culture flask containing 5 ml PYG medium. Wait for cell attachment.
14. Inoculate 50 µl of the previously collected well contents into the cell culture flask.
15. Incubate at 30°C for up to 7 days. Visually inspect cells daily to monitor for the appearance of CPE.
16. If CPE is confirmed, collect flask material, and store at –20°C for further identification. If CPE is not confirmed after 7 days, discard contents of the tissue culture flask.

Occasionally, false-positive results are observed. This is due to a cytotoxic effect (caused by some substance present in the sample that is toxic to cells), which causes nonspecific CPE. Because of this, the CPE confirmation procedure is important after collecting the well contents.

This procedure must be performed in a BSL-2 cabinet, and all materials must be sterile.

REAGENTS AND SOLUTIONS

PBS, 10×

Dissolve the following in 700 ml ultrapure water:
87.68 g NaCl
5.68 g Na_2HPO_4
2.72 g KH_2PO_4
Bring to 1 L with ultrapure water
Adjust pH to 7.2, if necessary, using 10 M NaOH or 4 M HCl
Autoclave in 1-L bottles for 30 min at 121°C, allowing bottles to cool at room temperature
Store at 4°C for up to 6 months

To make 1× PBS from 10× PBS, take 100 ml of 10× PBS and bring volume to 1 L with ultrapure water.

PYG medium

Combine PYG medium solutions 1 and 2 (see recipes) and mix thoroughly. Add ultrapure water to a final volume of 1 L. Adjust pH to 6.5, if necessary, using 10 M NaOH or 4 M HCl. Filter medium and then sterilize by autoclaving in 500-ml bottles for 20 min at 121°C. Allow bottles to cool at room temperature. In a BSL-2 cabinet, supplement medium with antibiotics and fungicides using 0.25 µg/ml amphotericin B, 100 Ul/ml penicillin, and 0.1 mg/ml streptomycin. Store at 4°C for up to 6 months.

PYG medium, solution 1

Dissolve the following in 300 ml ultrapure water:
0.98 g $MgSO_4 \cdot 7H_2O$
0.06 g $CaCl_2$
9.0 g glucose ($C_6H_{12}O_6$)
0.02 g ammonium iron (II) sulfate hexahydrate ($Fe(NH_4)_2(SO_4) \cdot 6H_2O$)
0.40 g $Na_2HPO_4 \cdot 7H_2O$
0.34 g KH_2PO_4
1.00 g sodium citrate tribasic dihydrate ($C_6H_5Na_3O_7 \cdot 2H_2O$)

PYG medium, solution 2

Dissolve the following in 300 ml ultrapure water:
20.0 g Bacto-peptone extract in 200 ml ultrapure water
2.0 g yeast extract in 200 ml ultrapure water

COMMENTARY

Background Information

Since the isolation of APMV, different techniques have been tested for the isolation of this virus that infects amoebas. In 2011, a technique was assessed that included filtering samples through a 1.2-µm filter (Arslan, Legendre, Seltzer, Abergel, & Claverie, 2011). Currently we know that this technique is not suitable for isolating giant viruses, as it only results in the isolation of viruses that manage to pass through these filters. Larger viruses, such as cedratvirus and pandoravirus, would still be retained on the filter and therefore not isolated.

In another technique, petri dishes containing agar were used. Amoebas were split on an agar surface forming a monolayer, and then samples were inoculated onto this cell monolayer. The objective of this technique was to identify the isolation of a virus from the observation of formation of lysis plaques (Gaia et al., 2013). However, this method frequently presents fungal contamination and needs to be validated with several viral groups.

In contrast, the direct inoculation technique presented here has proven to be efficient for the isolation of different giant viral groups. Furthermore, use of a 96-well microplate allows the processing of several samples at the same time. It is important to mention that this technique is better suited for low-mobility protists and must be adapted for high-mobility protists. There is also a potential for culture contamination, depending on the nature of the samples. Nonetheless, some preventive measures are considered to reduce or prevent these contaminations, including the use of antibiotics as described in Basic Protocol 2.

Critical Parameters and Troubleshooting

Due to the nature of the samples, contamination with fungi and bacteria may occur. Contamination it is more likely to occur within the first passage; therefore, the measures below were established to minimize or avoid these contaminations:

1. Supplement PYG medium with other antibiotics, in addition to those commonly used (herein referred as “additional antibiotics”).

2. Establish a dilution of the samples (1:10), which lowers the probability of contamination. On the other hand, the chance of isolating a virus in an aliquot that was not diluted is higher. Considering this, in the first passage for prospection, the original aliquot is inoculated in one well and the diluted aliquot in another well.

3. Perform three freeze-thaw cycles of the aliquots prior to processing in an attempt to promote lysis of the cells in the environmental samples. As giant virus particles are very stable, they can resist and remain viable after this procedure.

Despite all precautions taken, contamination can still occur. In this scenario, other alternatives can be established. When contamination of only one of the wells of the same aliquot occurs, the next passage must proceed with only the aliquot that was not contaminated. When contamination of two wells of the same aliquot occurs, it is necessary to freeze and thaw the original aliquot three more times, perform a new dilution (1:10), and further dilute to 1:100. Thus, for the next passage, only the 1:10 and 1:100 dilutions will be inoculated in order to avoid further contamination.

Understanding Results

This protocol enables the isolation of giant viruses that have *A. castellanii* as a natural host or at least as a permissive host. As these viruses generally cause similar CPE, it is important to check the cells daily for roundness and cell lysis while passaging. In addition, it is important to be aware of other cellular changes. It is more common for CPE to appear on the second or third passage.

Last, even when observing CPE in the well, it is necessary to carry out the confirmation process, as it may be a cytotoxic effect caused by some other component of the sample itself.

Time Considerations

There is no specific period of time to collect environmental samples, and it may vary with the purpose of the study, the amount of sample available, or the type of sample to be collected (see Basic Protocol 1).

The time required for sample preparation (Basic Protocol step 12) might vary depending on the type and number of samples that will be processed. Typically, it takes 15 to 30 min to aliquot and dilute each sample. Mud and soil samples require an additional 24 hr for initial sediment settling.

The amoeba cultivation process is continuous and should be done at least three times a week. It usually takes 40 to 60 min for cell splitting and preparation of the 96-well microplates for prospecting (Basic Protocol 2 steps 2 through 8).

Inoculation of the original and diluted aliquots into the 96-well microplates (Basic Protocol 2 steps 9 through 12) usually takes 30 min to be completed. Plaque observation for CPE should be done daily. Three passages must be carried out, with each passage lasting 7 days. For confirmation of CPE, it may take up to 7 days if there is no confirmation before these 7 days.

Acknowledgments

The authors thank the Laboratório de Vírus of Universidade Federal de Minas Gerais for the support provided. The authors also thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Pró-Reitoria de Pesquisa and Pró-Reitoria de Pós-Graduação of UFMG, and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for the financial support, which were fundamental for the development of these protocols. JSA is a CNPq researcher.

Author Contributions

Talita Bastos Machado: formal analysis, investigation, methodology, writing—original draft; **Isabella Luiza Martins de Aquino**: formal analysis, writing—original draft; **Jônatas Santos Abrahão**: conceptualization, formal analysis, funding acquisition, writing—review and editing.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Literature Cited

Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., … La Scola, B. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nature Communications*, *9*, 749. doi: 10.1038/s41467-018-03168-1.

Andrade, A. C., Arantes, T. S., Rodrigues, R. A., Machado, T. B., Dornas, F. P., Landell, M. F., … Abrahão, J. S. (2018). Ubiquitous giants: A plethora of giant viruses found in Brazil and Antarctica. *Virology Journal*, *15*, 22. doi: 10.1186/s12985-018-0930-x.

Andreani, J., Aherfi, S., Bou Khalil, J. Y., Di Pinto, F., Bitam, I., Raoult, D., … La Scola, B. (2016). Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. *Viruses*, *8*(11), 300. doi: 10.3390/v8110300.

Andreani, J., Khalil, J. Y. B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., … La Scola, B. (2017). Pacmanvirus, a new giant icosahedral virus at the crossroads between Asfarviridae and Faustoviruses. *Journal of Virology*, *91*(14), e00212–17. doi: 10.1128/JVI.00212-17.

Andreani, J., Khalil, J. Y., Baptiste, E., Hasni, I., Michelle, C., Raoult, D., … La Scola, B. (2018). Orpheovirus IHUMI-LCC2: A new virus among the giant viruses. *Frontiers in Microbiology*, *8*, 2643. doi: 10.3389/fmicb.2017.02643.

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., & Claverie, J. M. (2011). Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(42), 17486–17491. doi: 10.1073/pnas.1110889108.

Bajrai, L. H., Benamar, S., Azhar, E. I., Robert, C., Levasseur, A., Raoult, D., & La Scola, B. (2016). Kaumoebavirus, a new virus that clusters with Faustoviruses and Asfarviridae. *Viruses*, *8*(11), 278. doi: 10.3390/v8110278.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., … Raoult, D. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21848–21853. doi: 10.1073/pnas.0911354106.

de Souza, G. A. P., Queiroz, V. F., Lima, M. T., de Sousa Reis, E. V., Coelho, L. F. L., & Abrahão, J. S. (2020). Virus goes viral: An educational kit for virology classes. *Virology Journal*, *17*(1), 13. doi: 10.1186/s12985-020-1291-9.

Gaia, M., Pagnier, I., Campocasso, A., Fournous, G., Raoult, D., & La Scola, B. (2013). Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter. *PLoS One*, *8*(4), e61912. doi: 10.1371/journal.pone.0061912.

Kerepesi, C., & Grolmusz, V. (2016). Giant viruses of the Kutch Desert. *Archives of Virology*, *161*(3), 721–724. doi: 10.1007/s00705-015-2720-8.

Kerepesi, C., & Grolmusz, V. (2017). The "Giant Virus Finder" discovers an abundance of giant viruses in the Antarctic dry valleys. *Archives of Virology*, *162*(6), 1671–1676. doi: 10.1007/s00705-017-3286-4.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., … Raoult, D. (2003). A giant virus in amoebae. *Science*, *299*(5615), 2033–2033. doi: 10.1126/science.1081867.

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., … Claverie, J. M. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(11), 4274–4279. doi: 10.1073/pnas.1320670111.

Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., … Claverie, J. M. (2015). In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(38), E5327–E5335. doi: 10.1073/pnas.1510795112.

Mihara, T., Koyano, H., Hingamp, P., Grimsley, N., Goto, S., & Ogata, H. (2018). Taxon richness of "Megaviridae" exceeds those of bacteria and archaea in the ocean. *Microbes and Environments*, *33*(2), 162–171. doi: 10.1264/jsme2.ME17203.

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., … Abergel, C. (2013). Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science*, *341*(6143), 281–286. doi: 10.1126/science.1239181.

Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., … La Scola, B. (2015). Faustovirus, an Asfarvirus-related new lineage of giant viruses infecting amoebae. *Journal of Virology*, *89*(13), 6585–6594. doi: 10.1128/JVI.00115-15.

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., … Woyke, T. (2017). Giant viruses with an expanded complement of translation system components. *Science*, *356*(6333), 82–85. doi: 10.1126/science.aal4657.

Schulz, F., Alteio, L., Goudeau, D., Ryan, E. M., Yu, F. B., Malmstrom, R. R., … Woyke, T. (2018). Hidden diversity of soil giant viruses. *Nature Communications*, *9*(1), 4881. doi: 10.1038/s41467-018-07335-2.

Wu, S., Zhou, L., Zhou, Y., Wang, H., Xiao, J., Yan, S., & Wang, Y. (2020). Diverse and unique viruses discovered in the surface water of the East China Sea. *BMC Genomics*, *21*(1), 441. doi: 10.1186/s12864-020-06861-y.

Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., … Takemura, M. (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *Journal of Virology*, *93*(8), e02130–18. doi: 10.1128/JVI.02130-18.

4.2 – Artigo 2 – Gene duplication as a major force driving the genome expansion in some giant viruses

Este artigo foi publicado no periódico Journal of Virology.

Vírus gigantes com seus genomas gigantescos estão entre os componentes mais intrigantes da virosfera. Como esses vírus adquiriram tais genomas gigantes permanece obscuro, apesar dos esforços consideráveis para entender esse fenômeno. Aqui, descrevemos a descoberta do cedratvirus pambiensis, um vírus gigante de ameba isolado no Brasil. Embora a morfologia da partícula e o ciclo de replicação do c. pambiensis sejam muito semelhantes aos descritos para outros cedratvirus, o sequenciamento do genoma completo revelou o maior genoma de cedratvirus já descrito, com 623.564 pares de bases e 842 genes codificadores de proteínas previstos (entre eles, 76 ORFans). A análise do genoma revelou um número sem precedentes de genes parálogos, com ~73% do genoma do c. pambiensis sendo composto de genes com duas ou mais cópias. Grandes famílias de genes parálogos funcionalmente diversos incluíam até >70 cópias e foram distribuídas por todo o genoma. A investigação aprofundada dos mecanismos e origens das duplicações genéticas revelou que tanto as duplicações em tandem quanto a transferência distal de blocos sintênicos de genes contribuíram para a expansão genômica do c. pambiensis. Finalmente, uma análise abrangente do genoma de vírus de todas as famílias de vírus gigantes conhecidas sugeriu que a duplicação genética é um dos principais mecanismos subjacentes ao gigantismo genômico em todo o filo *Nucleocytoviricota*. A expansão dos genomas virais por meio de duplicações sucessivas seguidas pela subfuncionalização e exaptação das cópias gênicas parálogas pode promover a adaptação de vírus gigantes a uma variedade de nichos.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Virology | Full-Length Text

Gene duplication as a major force driving the genome expansion in some giant viruses

Talita B. Machado,[1] Agnello C. R. Picorelli,[2] Bruna L. de Azevedo,[1] Isabella L. M. de Aquino,[1] Victória F. Queiroz,[1] Rodrigo A. L. Rodrigues,[1] João Pessoa Araújo Jr.,[3] Leila S. Ullmann,[3] Thiago M. dos Santos,[4] Rafael E. Marques,[5] Samuel L. Guimarães,[5] Ana Cláudia S. P. Andrade,[6] Juliana S. Gularte,[7] Meriane Demoliner,[7] Micheli Filippi,[7] Vyctoria M. A. G. Pereira,[7] Fernando R. Spilki,[7] Mart Krupovic,[8] Frank O. Aylward,[9,10] Luiz-Eduardo Del-Bem,[4] Jônatas S. Abrahão[1]

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT Giant viruses with their gigantic genomes are among the most intriguing components of the virosphere. How these viruses attained such giant genomes remains unclear, despite considerable efforts to understand this phenomenon. Here, we describe the discovery of cedratvirus pambiensis, an amoebal giant virus isolated in Brazil. Although the virion morphology and replication cycle of c. pambiensis are very similar to those described for other cedratviruses, whole genome sequencing revealed the largest cedratvirus genome ever described, with 623,564 base pairs and 842 predicted protein-coding genes (among them, 76 ORFans). Genome analysis has revealed an unprecedented number of paralogous genes, with ~73% of the c. pambiensis genome being composed of genes with two or more copies. Large families of functionally diverse paralogous genes included up to >70 copies and were distributed across the genome. The in-depth investigation of the mechanisms and origins of gene duplications revealed that both tandem-like duplications and distal transfer of syntenic blocks of genes contributed to the c. pambiensis genomic expansion. Finally, a comprehensive genome analysis of viruses from all known giant virus families suggested that gene duplication is one of the key mechanisms underlying genomic gigantism across the phylum *Nucleocytoviricota*. The expansion of viral genomes through successive duplications followed by subfunctionalization and exaptation of the paralogous gene copies may promote the adaptation of giant viruses to a variety of niches.

IMPORTANCE Giant viruses are noteworthy not only due to their enormous particles but also because of their gigantic genomes. In this context, a fundamental question has persisted: how did these genomes evolve? Here we present the discovery of cedratvirus pambiensis, featuring the largest genome ever described for a cedratvirus. Our data suggest that the larger size of the genome can be attributed to an unprecedented number of duplicated genes. Further investigation of this phenomenon in other viruses has illuminated gene duplication as a key evolutionary mechanism driving genome expansion in diverse giant viruses. Although gene duplication has been described as a recurrent event in cellular organisms, our data highlights its potential as a pivotal event in the evolution of gigantic viral genomes.

KEYWORDS giant virus, *Pithoviridae*, cedratvirus pambiensis, genome expansion, paralogous genes, *Nucleocytoviricota*

Gene and genomic segment duplication is a critical mechanism underlying the evolution of cellular organisms by providing raw genetic material for the emergence of new gene functions and pathways (1). Duplicated genes can undergo subfunctionalization by acquiring mutations, resulting in the evolution of new protein

Editor Colin R. Parrish, Cornell University Baker Institute for Animal Health, Ithaca, New York, USA

Address correspondence to Jônatas S. Abrahão, jonatas.abrahao@gmail.com, or Luiz-Eduardo Del-Bem, delbem@ufmg.br.

The authors declare no conflict of interest.

See the funding table on p. 15.

Received 23 August 2023
Accepted 26 October 2023
Published 21 November 2023

functions. The significance of this phenomenon in evolution is evidenced by the widespread occurrence of duplicated genes across all domains of life (2–5). It has been estimated that around 30%–65% of the genes in multicellular eukaryotes, such as humans, have emerged through duplication (6, 7).

Giant viruses of amoeba are characterized by their large genome sizes and complex gene repertoires (8–11). Although the driving forces that led to the genome gigantism of those viruses are not fully understood, horizontal gene transfer, *de novo* gene emergence, and gene duplication have all been hypothesized to have contributed to genome expansion (12–18). Large families of functionally diverse paralogous genes have been identified in giant viruses of amoebas, including genes encoding ankyrin repeat-containing proteins, receptors for ubiquitination targets, and proteins with glycosyltransferase domains. In addition, many of these gene families are composed of unknown proteins or ORFan genes (13, 19). Although studies on gene duplication in giant viruses are scarce, there is convincing evidence showing that approximately one-third of the mimivirus genome and 50% of pandoravirus genomes are composed of multi-copy genes (13, 19).

Here, we report the discovery of cedratvirus pambiensis, a giant amoeba virus with the largest genome size ever described for the cedratvirus group, comprising 623,564 base pairs. The investigation of the architecture of the genome revealed an unprecedented abundance of duplicated genes, which constitute up to 72% of the total genome, a proportion on par with or surpassing that observed in cellular organisms. Most of these genes are grouped into six major gene families. Only 27.7% of the genome is composed of single-copy genes (non-duplicated genes). The expansion of the analyses to other varidnaviruses revealed extensive gene duplications in most groups of giant viruses, most markedly in members of the "Pithoviridae"-related viruses (which includes cedratviruses), pandoraviruses, and some mimiviruses. The discovery of c. pambiensis expands our understanding of the diversity and complexity of giant viruses, emphasizing the role of gene duplication in driving their genome expansion and shaping the genomic content.

RESULTS

Cedratvirus pambiensis particles and replication cycle

As part of our ongoing efforts to characterize the diversity of giant viruses infecting amoeba, we have isolated a new cedratvirus, c. pambiensis, from a water sample collected in a small, forested area at the Federal University of Minas Gerais (UFMG) campus, Belo Horizonte, Brazil. Inoculated amoebas exhibited cytopathic effects such as rounding and lysis, and upon light microscopy examination, it was possible to visualize viral particles. To gain a more comprehensive understanding of the particles' characteristics, we analyzed images obtained by transmission electron microscopy (TEM), negative staining electron microscopy (NSEM) (Fig. 1), and scanning electron microscopy (SEM) (see Supplementary Fig. 1a posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses). The virus particles were oval-shaped, measuring approximately 1 µm in length and 500 nm in width (Fig. 1a). The capsid is composed of parallel striations (Fig. 1a) and may have one or two apical "corks." Most observed particles had two corks, which is consistent with previous descriptions of cedratviruses. Notably, the capsid of c. pambiensis exhibited surface fibrils, a structure that has not been previously documented in cedratviruses (Fig. 1c). Although not visible by TEM and SEM, these fibrils were present in all images obtained by NSEM.

The replication cycle of c. pambiensis is a complex process that involves several steps (Fig. 1). It begins with the entry of viral particles into the host cell, which probably occurs via phagocytosis. Once inside the cell, the viral particles undergo uncoating, followed by an eclipse period which typically lasts for around 3 hours (Supplementary Fig. 1b posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses). The uncoating

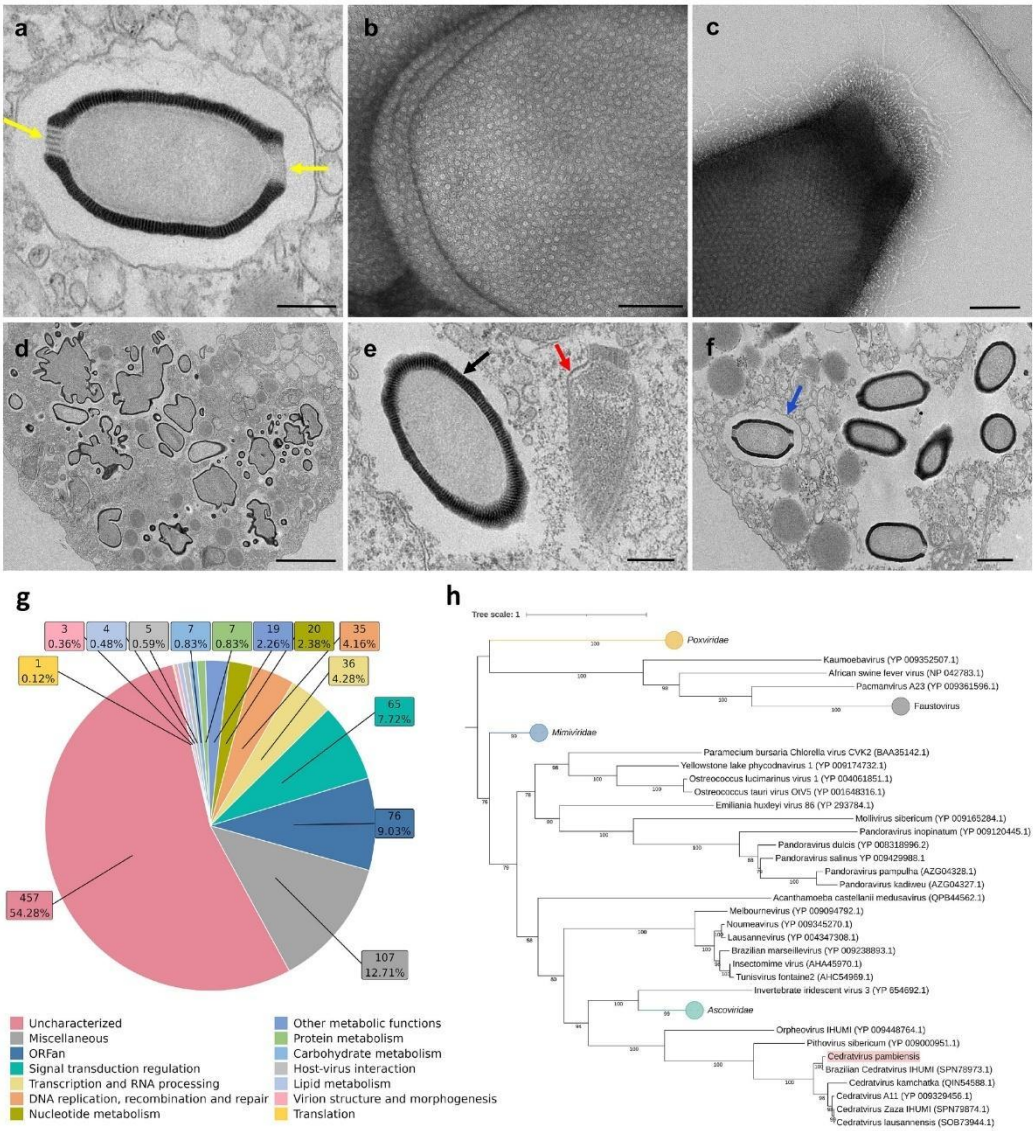

FIG 1 Particle, replication cycle, and genomic features of c. pambiensis. (a) Cedratvirus particle showing an oval shape with two apical corks (yellow arrows). Scale bar: 200 nm. (b) Zoom in the particle, showing the parallel striated structures of the capsid. Scale bar: 100 nm. (c) Observation of surface fibrils on the capsid. Scale bar: 100 nm. (d) Amorphous structures with no defined function in the cytoplasm of an infected amoeba. Scale bar: 2 µm. (e) Assembly of new viral progeny. It is possible to observe structures of early particles being formed (red arrow) and mature particles (black arrow) at the same time inside the cells. Scale bar: 200 nm. (f) A viral particle was observed inside a vesicle after assembly (blue arrow), suggesting release by exocytosis before cell lysis. Scale bar: 500 nm. (g) Functional categories of c. pambiensis predicted genes. The color legend is provided below the graph. (h) Maximum likelihood phylogenetic tree constructed with amino acid sequences from the DNA polymerase subunit B of cedratviruses and other nucleocytoviruses. The new isolate described here, c. pambiensis (highlighted in pink), clustered with other cedratviruses, closer to the Brazilian cedratvirus. The scale bar indicates the genetic distance.

involves the removal of the viral cork to release the viral genome into the cytoplasm of the host cell. The next step is the formation of the viral factory, which is a region within the host cell that supports viral replication. In electron micrographs, the viral factory appears as a space that is unbounded and electron-lucent. Within the factory, amorphous electron-dense structures can be observed (Fig. 1d). Although the function of

TABLE 1 Comparison between the genomes of all cedratviruses with published genomes

Virus	Genome size	Predicted proteins	% Intergenic regions
Cedratvirus A11	589,068 bp	574	18.99
Cedratvirus lausannensis	575,161 bp	643	17.15
Cedratvirus zaza	560,887 bp	636	15.69
Brazilian cedratvirus	460,038 bp	533	14.43
Cedratvirus kamchatka	466,767 bp	545	18.26
Cedratvirus pambiensis	**623,564 bp**	**842**	**10.59**

these structures is unknown, they appear to be composed of material that is similar to that of the viral capsid, as seen in Fig. 1a. During the assembly of new viral particles, initial structures that will contribute to the formation of new virions can be observed, as shown in Fig. 1e. At 12 hours post-infection (hpi), viral production reaches its maximum level, after which it plateaus, as seen in Supplementary Fig. 1b posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses. The final step in the replication cycle is the release of viral progeny, which occurs via cell lysis. However, exocytosis may also play a role in this process, as depicted in Fig. 1f, where particles can be visualized inside vesicles close to the cell plasma membrane at the end of the infection cycle.

Cedratvirus pambiensis has an unprecedented abundance of paralogous genes

Genomic characterization of c. pambiensis yielded a circular dsDNA molecule of 623,564 bp and encoding 842 predicted proteins. Until then, the largest cedratvirus genome was described for cedratvirus A11, with 589,068 bp, and the largest number of predicted proteins was described for cedratvirus lausannensis, with 643. Thus, c. pambiensis is the cedratvirus with the largest genome and highest number of predicted proteins among all cedratviruses published to date (Table 1). As gene prediction methods may vary among different studies, we performed the gene prediction of all available cedratviruses using the same parameters that we applied to c. pambiensis. Although we observed a general increase in the number of predicted genes for all viruses, c. pambiensis still holds the record for the largest number of predicted genes among the isolated viruses. Functional analysis of the predicted proteins (Fig. 1g) revealed that most of them are uncharacterized (54.24%), and ORFans (9.03%). Proteins related to the regulation of signal transduction; transcription and RNA processing; DNA replication, recombination and repair, and different types of metabolism were also identified. The construction of a phylogenetic tree using amino acid sequences from the family B DNA polymerase showed that the new isolate clustered with other cedratviruses (Fig. 1h), and most closely to the Brazilian cedratvirus. The synteny analysis reinforces the proximity of these two cedratviruses, when compared to the other cedratviruses (see Supplementary Fig. 2 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses).

Initially, we hypothesized that the increase in the genome size could be due to the presence of a new class of genes or a substantial number of ORFans. However, the annotation of the c. pambiensis genome revealed a gene content that is similar to that of other cedratviruses. Next, we investigated the intergenic content of c. pambiensis genome and compared it to that of other cedratviruses. However, after analyzing the intergenic content in all cedratviruses with available genomes, we observed that c. pambiensis has the lowest percentage (10.59%) of predicted intergenic regions among all of them (mean of 16.90%) (see Supplementary Table 1 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses).

Then, we tested whether the larger genome size of c. pambiensis could be explained by the increase in the number of paralogous genes, which occurs due to gene and genomic segment duplications. We performed an all-against-all BLASTp analysis of the predicted proteins of c. pambiensis, which revealed an unexpectedly large number of paralogous groups with more than three genes, as well as paralogs grouped in pairs and triplets (Fig. 2). Only 27.7% of the c. pambiensis genome consists of single copy genes (Fig. 2). At least six large gene families (>20 genes) were identified in c. pambiensis genome, encoding functionally diverse proteins, including ankyrin-domain containing proteins, collagen-like proteins, serine-threonine protein kinases, hypothetical proteins, proteins containing F-box domain, ORFans, and others. Certain families presented genes with more than one predicted function or domain, suggesting progressive differentiation after duplication (Fig. 2). All the information about the paralogous genes is described in Supplementary Material 1 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses.

Of 842 total genes, 609 (72.3%) are part of multi-gene families while only 233 (27.7%) are single-copy genes (Fig. 3a). We observed that a large fraction of these paralogous groups (34.6%) are composed of tandem genomic segment duplications (Fig. 3a and b). We then evaluated the percentage of tandemly duplicated genes across gene families, showing that these events are much more common in the six largest gene clusters (57%), while in low-copy number families present in triplets (9%) or pairs of genes (10%) tandem duplication events are less frequent (Fig. 3c). This finding suggests that tandem duplications are primarily responsible for the formation and expansion of large gene families.

To better understand the tandem duplication events and their effect on the genome evolution of these viruses, we constructed phylogenetic trees with the protein sequences for the six largest gene families (or clusters). Analysis of the phylogenetic trees showed that two types of duplication events occurred: proximal tandem duplications (involving more recent duplications) and chromosomal segment duplications (when an entire block of tandem genes appears to have been copied from one part of the genome and pasted into another, occasionally disrupting the preexisting genes). As an example, these analyses were detailed for the phylogenetic tree of cluster 3 (Fig. 3d), in which tandem genes were identified (Fig. 3e). Both proximal tandem duplication events (Fig. 3f) and chromosome segment duplication events (Fig. 3g) can be observed, and we note that these events can follow each other. Taking the tandem genes of groups A (207-208-209-210-211) and B (241-242-243-244) as an example, two interpretations can be made: (i) a copy-paste event occurred from group A to group B, and a gene was lost afterward; (ii) there was a copy-paste event from group B to group A, and subsequently, a proximal tandem duplication gave rise to gene 208. The phylogenetic trees for the other large gene families can be seen in Supplementary Fig. 3 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses. We quantified the two events (proximal tandem duplications and chromosomal segment duplications) for the six largest gene families (see Supplementary Fig. 4 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses) and noticed that they are frequent and seem to have a great influence on the genome evolution of this virus.

After a gene duplication event, the copies follow different evolutionary paths. When one of the copies suffers an extreme reduction in its coding sequence (CDS) caused by the emergence of a premature stop codon, we can infer that a pseudogenization event has occurred (resulting in a progressive loss of gene function). Our data suggest that there was a significant difference in CDS length and identity among genes within the

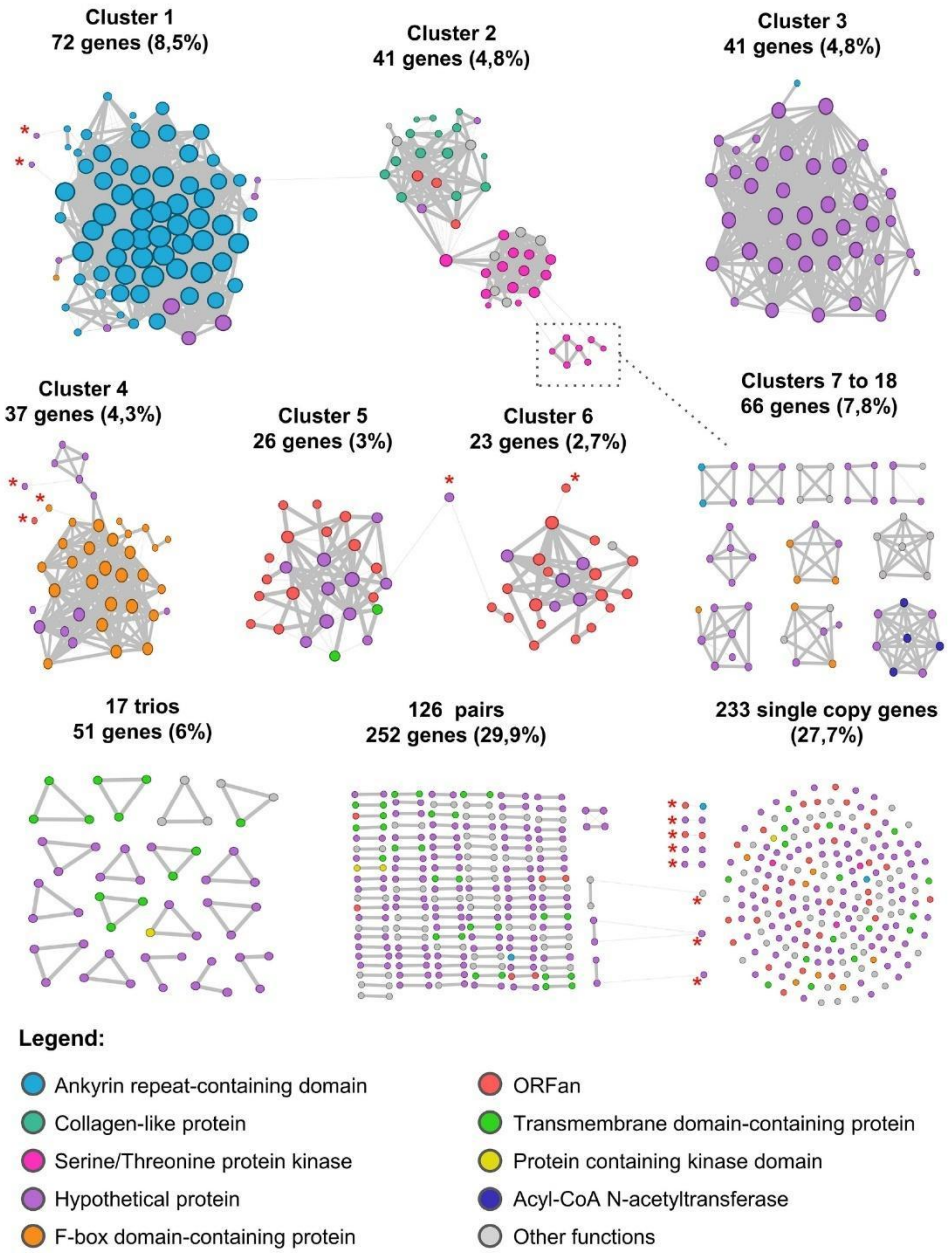

FIG 2 Network of clusters, trios, pairs, and single copy genes in the c. pambiensis genome. Reciprocal BLASTp-hits (coverage ≥30 and e-value < 1e−-4) between two proteins are represented by thicker lines, while non-reciprocal BLASTp-hits are represented by thinner lines. Considering the defined criteria, to be part of the cluster, the gene must: (1) have a reciprocal match with some gene within the cluster (thick line) or (2) have at least two non-reciprocal matches with genes within the cluster (thin line). Genes that only had one non-reciprocal match with a gene within the cluster were not considered part of that cluster. Asterisks indicate single genes, which were not included within the clusters according to the aforementioned criteria. The color legend is provided below the image. The square highlights cluster 10, in which has non-reciprocal hits with cluster 2.

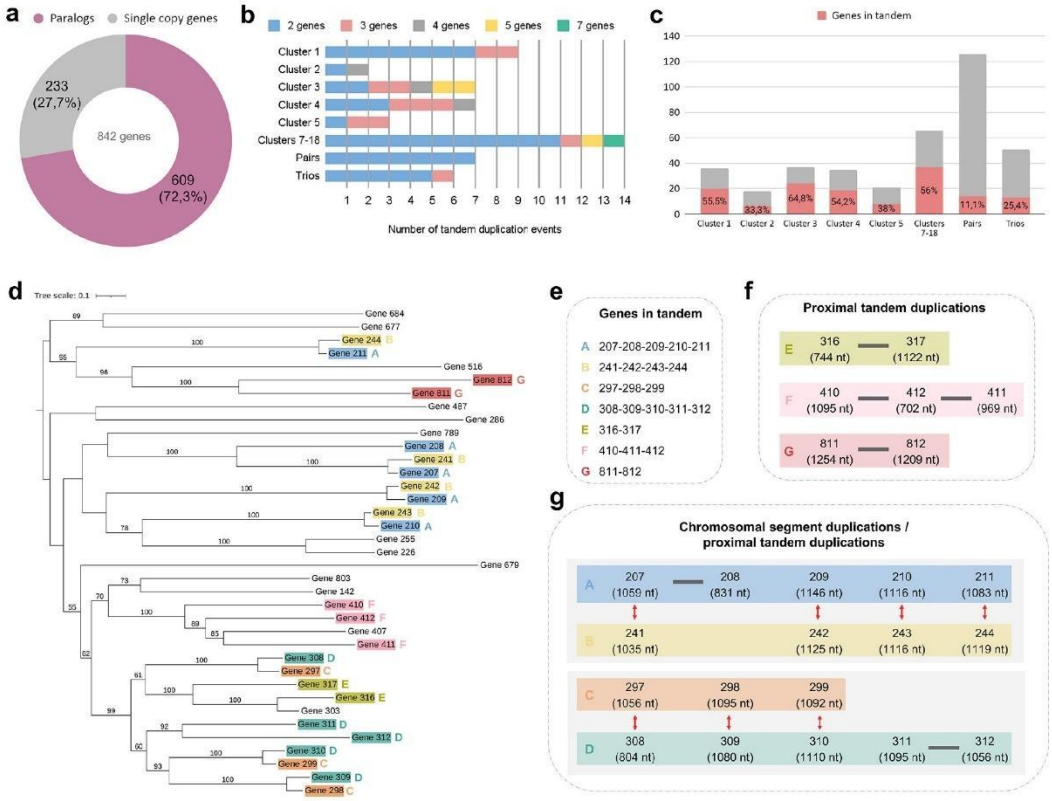

FIG 3 Gene duplication analyses in the c. pambiensis genome. (a) Number and percentage of genes in paralog groups or single copy genes in c. pambiensis genome. (b) Number of tandem duplication events observed within paralog clusters with more than three genes and low-copy number clusters composed of trios and pairs of genes. (c) Percentage of tandem duplication events observed within paralogs clusters, trios, and pairs. (d) Maximum likelihood phylogenetic tree constructed with protein sequences encoded by cluster 3 genes. Tandem genes are highlighted and colored in the tree according to the organization shown in (e). (e) Tandem genes highlighted in (d). They were organized in ascending order. (f) Proximal tandem duplication events are identified in (d). (g) Chromosomal segment duplications plus proximal duplication events identified in (d). In (f) and (g), the length of the gene in nucleotides is depicted below the gene ID.

same family, indicating non-negligible pseudogenization (Fig. 4a, and see Supplementary Fig. 5 and 6 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses). In addition to gene size/coverage variation, we also observed considerable sequence divergence of the paralogous proteins, so that some of the cluster members were not reciprocally identifiable as homologous in BLASTp analysis, suggesting independent and progressive evolution after gene duplication (Fig. 3 and 4, and see Supplementary Fig. 5 and 6 posted at https://www.giantviruses.com/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses).

The chromosome position of the paralogs can provide clues about how genome expansion has evolved. We observed that some clusters (1, 3, and 4) appear to have more gene copies concentrated in a given region within the genome, while others (2, 5, and 6) are more spread throughout the genome (Fig. 4b). But in general, the paralogs belonging to those six major gene families are scattered throughout the c. pambiensis genome (see Supplementary Fig. 7 posted at https://www.giantviruses.com/sup-material-of-papers/

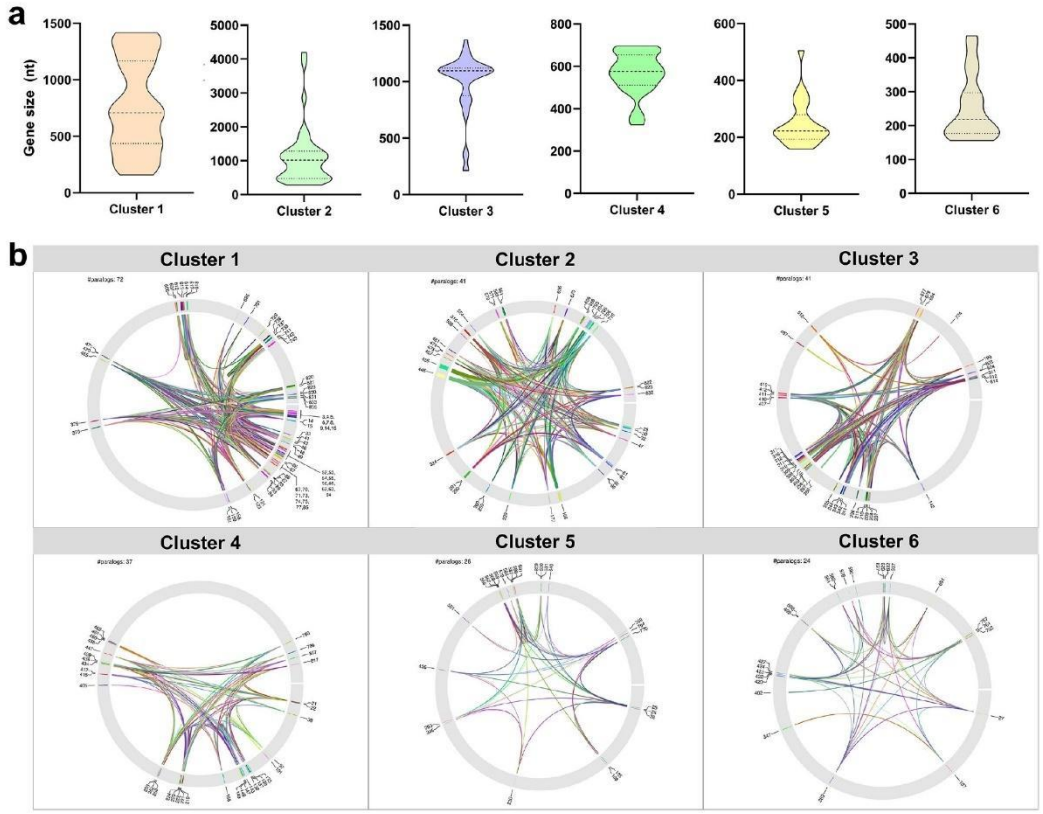

FIG 4 Size and location of paralogous genes comprising each large family of genes (clusters 1 to 6). (a) Violin plots showing gene size variation considering each cluster. Coding sequences far below the mean may suggest possible pseudogenization events. The dashed line represents the mean. Dotted lines delimit the interquartile range. (b). Gene location considering each cluster. Clusters 1, 3, and 4 seem to have some polarity in the genome, while clusters 2, 5, and 6 do not seem to have any pattern and are scattered throughout the genome.

sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses), indicating multiple and successive events of gene/chromosome segment duplication.

Gene duplication as a driving force of genome gigantism in giant viruses

To explore the prevalence and distribution of paralogs in cedratviruses, we expanded our analysis to include all available cedratvirus genomes in public databases (Fig. 5). By performing BLASTp searches of predicted proteins against the complete set of proteins of each cedratvirus, we identified a range of gene families with varying sizes and predicted functions. Firstly, it is noteworthy that c. pambiensis has an atypical and unprecedented relative abundance of duplicated genes compared to other cedratviruses, accounting for 72.3% of its genome. Nevertheless, a significant contribution of paralogs was observed in all cedratvirus genomes, ranging from 43.34% in Brazilian cedratvirus to 52.26% in c. lausannensis. In addition, all cedratviruses share some large gene families, such as families mainly related to ankyrin repeat-containing domain and hypothetical proteins, as well as other functions such as collagen-like proteins, serine/threonine kinases and F-box domain-containing proteins. We further expanded our analysis to include the pithovirus-like group, to which cedratviruses belong. Interestingly, pithovirus and orpheovirus present a similar proportion of duplicated genes in their genomes, at 42.61% and 52.29%, respectively, but other protein domains were

overrepresented, such as collagen-like and MORN-repeat proteins, suggesting that extensive duplications have occurred independently following the radiation of the pithovirus-like group.

Expanding the analysis to other members of the phylum *Nucleocytoviricota* and yaravirus revealed that gene duplications are, again, quite abundant (Fig. 6a). In addition to cedratviruses, gene duplication also seems to be an important factor in the evolution of the genome of other giant viruses, particularly for pandoraviruses (mean of 47.28%) and some mimiviruses (e.g., 49.46% for cotonvirus). Despite the different proportions, all analyzed genomes have duplicated genes. However, it is important to highlight that even after expanding our analysis to other nucleocytoviruses, c. pambiensis remains the virus with the highest percentage of the genome composed of duplicated genes (72.3%). We note, however, that for some viruses, mechanisms other than gene duplication may be acting synergistically or concurrently.

It is generally expected to observe a positive correlation between viral genome size and the number of genes. The isolation of c. pambiensis raised questions about this correlation and the existence of a correlation between genome size and the number of paralogs. To address these questions, we compared genome size with the overall number of genes and the number of duplicated genes across a large sample of giant viruses (Fig. 6b and c). As aforementioned, variations on gene prediction methods must be considered, but the overall available data strongly suggest the presence of a strong positive correlation between genome size and both the total number of genes ($\rho = 0.90$, *P*-value $< 2.2e{-}16$) and the number of paralogs ($\rho = 0.87$, *P*-value = 1.4e−-14) per genome. The linear regression analysis reveals that, although all cedratviruses show a similar correlation between genome size and the number of predicted genes, c. pambiensis stands out as an exceptional case due to the significant contribution of paralogs in its genome (Fig. 6c).

DISCUSSION

The genome gigantism observed in giant viruses represents an intriguing unanswered question. A number of giant viruses have been discovered in recent years, revealing an increasing variety of particles, genome sizes, and predicted genes. Although deserving attention has been given to the functional content of the giant virus genomes, few studies have investigated why the genomes of these viruses are so large, reaching up to 2.8 Mb (8–11, 14, 16). The early efforts to answer this question were hampered by the scarcity of genomic information available at the time, precluding generalizing conclusions. However, the constant efforts of several research teams to isolate novel giant viruses worldwide have now set the stage for a more comprehensive analysis of giant virus genome evolution. Here, we presented the evidence that gene duplication is a primary mechanism for genome expansion among several groups of giant viruses.

Gene duplication has been recognized as an important source of genetic diversity in cellular organisms (2–5). Through gene duplication, new functions can emerge, as duplicated genes typically experience lower negative selection pressures, and the encoded proteins can gain new properties and functions. Gene families within a given organism typically emerge as a result of duplication events followed by divergence (20). In addition to potential functional divergence, the multi-copy gene families can contribute to the increased gene expression via so-called gene dosage phenomenon. Some of the well-known examples of multi-gene families include genes encoding cytomotive filament-forming proteins, globins, and ribosomal units (20–22). As a result, gene duplication is consistently considered a major mechanism in the evolution of cellular organisms. However, the understanding of the consequences of gene duplication in viruses remains limited. There are ample examples showing that gene duplication followed by exaptation of one of the gene copies has played a key role in adaptation and diversification throughout virus evolution (23, 24). Nevertheless, in viruses with small capsids, experimental studies have shown that gene duplications are prohibitive and lead to the loss of infectivity, primarily dictated by the limited packaging capacity of

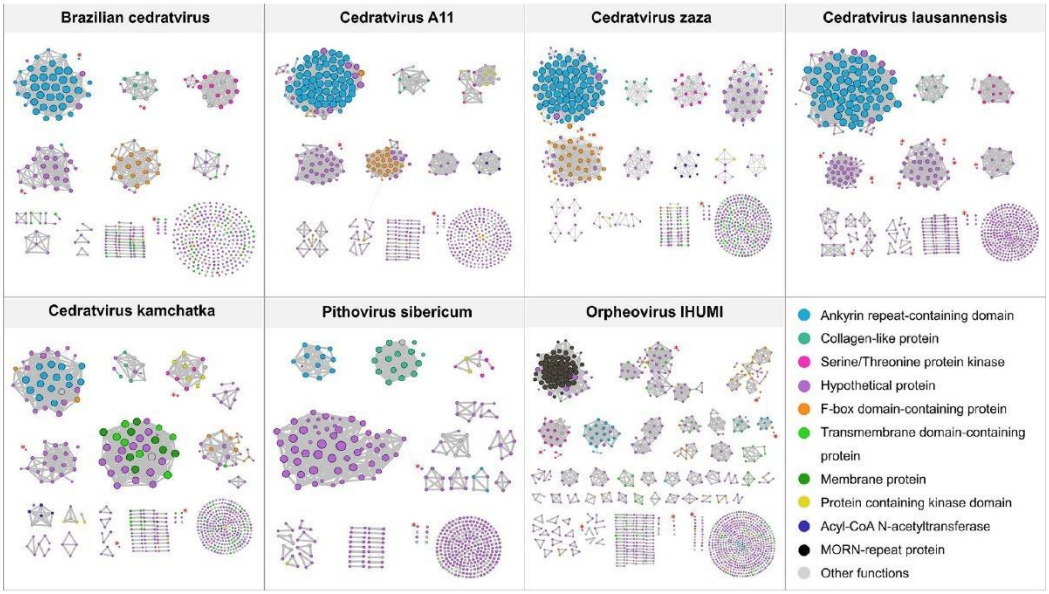

FIG 5 Comparison of gene families in the genome of members of the pithovirus-like group. These networks were made in the same way as for c. pambiensis (Fig. 2). Color legend is provided on the image.

small capsids (25). The situation in giant viruses, which show overall low packaging densities (26), is radically different compared to viruses with small capsids. Thus, genome evolution in giant viruses is apparently not constrained by the capsid size, allowing them to reap the benefits of gene duplication, which has shaped the genomes of cellular organisms.

In this study, we have described that genes in c. pambiensis appear to duplicate through several mechanisms, involving both tandem gene and distal genomic segment duplications. Tandem duplications in cellular organisms may occur through replication slippage, ectopic recombination, or aberrant DNA break repair. Distal duplications typically involve unequal crossing-over rearrangements of gene clusters with similar gene content. This mechanism may become progressively more frequent as the number of paralogs increases, providing more regions with similar content available for recombination (1). It is notable that the largest paralogous gene families encode proteins that themselves consist of repetitive domains, such as ankyrin repeats, leucine-rich repeats, and MORN repeats. Conceivably, the repetitive nucleotide sequences within these genes promote both tandem and long-distance genomic duplications. Furthermore, considering that a substantial number of viral genome copies are produced and compacted within viral factories, and considering the fact that the cedratvirus genome is circular dsDNA, it is reasonable to believe that ectopic recombination and unequal crossing-over may generate both tandem and distal duplications in cedratviruses. Additionally, the role of transposons should be considered in relation to gene duplication, as they have been described in the genomes of amoeba and certain groups of giant viruses (27, 28). Considering that the *Nucleocytoviricota* may have arisen from smaller and simpler viruses infecting early eukaryotes (29–31), the expansion of their genomes might involve a general mechanism conserved in the entire phylum. While genome expansion by extensive gene gain through horizontal gene transfer and *de novo* gene creation seem to be characteristic of only some groups of nucleocytoviruses (13, 15), our data indicate that gene duplication is an evolutionary mechanism common for several groups of giant viruses, especially those capable of infecting amoebas.

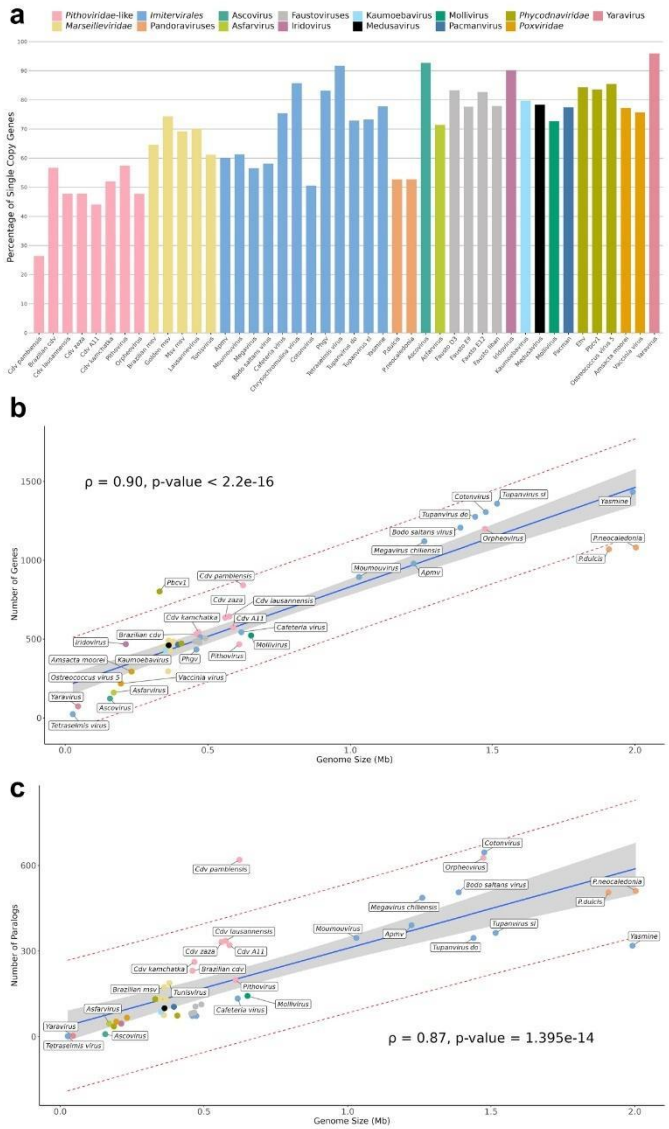

FIG 6 The contribution of paralogs genes in the genomes of giant viruses. (a) Comparison of the percentage of single copy genes in the genome for members of *Nucleocytoviricota* and yaravirus. The representatives of each group were depicted with distinct colors, as described in the legend. The percentage of single-copy genes is shown for each genome. Spearman correlation plots of the relationship between genome size and total number of predicted genes (b) and the relationship between genome size and total number of paralogs (c). Yasminevirus and c. pambiensis are outliers. The solid blue line marks the linear regression while the shaded gray area illustrates the 95% confidence interval associated with the linear regression line. The outer red lines delineate the 95% prediction interval, encapsulating the range within which we anticipate 95% of gene/paralog numbers to fall based on a given genome size. ρ: Spearman correlation coefficient; *p*-value: associated *P*-value.

One question that arises is why cedratviruses have so many duplicated genes in their genome. Maintaining duplicated copies may be important for creating genetic redundancy, and protecting the virus from deleterious mutations in essential genes since

additional copies could maintain the organism's functionality and fitness (32). Indeed, some of the c. pambiensis duplicated genes are potentially essential, such as the major capsid protein, early transcription factors, and transcriptional enzymes. Furthermore, as aforementioned, gene duplication is a phenomenon that provides raw material for evolution. The additional gene copies can be repurposed for functions unrelated to those of the original genes, which appears to be the main trend in virus evolution (23, 33). Furthermore, the duplicated genomic regions provide the raw genetic material for *de novo* gene emergence, a route extensively explored by pandoraviruses (13). Both mechanisms can lead to genetic innovation, increasing the genetic repertoire of these viruses. This is supported by the presence of different identifiable domains/functional categories in certain clusters of paralogous genes in c. pambiensis and other pitho-like viruses (Fig. 2 and 5). Further exploration of the giant virus diversity should further refine our understanding of the mechanisms of genome expansion and evolutionary traits associated with this remarkable group of viruses.

MATERIALS AND METHODS

Sample collection and isolation

To obtain the sample, a plastic container was placed in a small, forested area at the UFMG campus for a few days, collecting rainwater and organic matter present in that environment. From this collected water, the method of co-culture with amoebae of the species *Acanthamoeba castellanii* was carried out, as described in a published protocol (34). Collection authorization: SISBIO 89441-1. Brazilian genetic resources access authorization: SISGEN A2291C9.

Production, purification, and titration

To produce the new isolate, *A. castellanii* were infected with an MOI of 0.01 in glass culture flasks (300 cm^2) with 35 mL of PYG medium and kept at 30°C in a rotary cell oven. After complete lysis of the cells, the contents of the flask were collected. This content was added to a sucrose cushion (40%) and then ultracentrifuged in a Combisorvall Rotor AH-62va centrifuge at 14,000 rotations per minute (AH-629 rotor) for 1 hour, between 4°C and 8°C. The final pellet was resuspended in phosphate-buffered saline. Titration was performed by the limiting dilution method (35) in 96-well plates and the titer was expressed in $TCID_{50}$ per milliliter. The viral stock was kept at −20°C until use.

Electron microscopy and one-step growth curve

Three electron microscopy methods were used during this work, for a better description of the viral particle: TEM, negative contrast electron microscopy, and SEM. For TEM, *A. castellanii* cells cultivated in a PYG medium were infected with an MOI of 0.01 for 24 hours. Cells were then collected and fixed with 2.5% glutaraldehyde + phosphate for 2 hours at room temperature. Subsequently, fixation was performed with 2% osmium tetroxide, and incorporation in EPON resin, in sequence ultrathin sections was made. Image analyses were performed using a transmission electron microscope (FEI SpiritBiotwin 120 kV). For NSEM, the purified virus was diluted 1:10 in water, and 3 µL of this diluted sample was applied onto glow-discharged 400-mesh copper grids covered with a Lacey carbon support film and an ultrathin carbon layer (15 mA, negative charge for 40 seconds, 01824—Ted Pella, USA). After 1 minute, the excess liquid was drained gently touching the edge of the grid with a filter paper. The grid was stained twice with 3 µL uranyl acetate solution (2%) for 30 seconds. The excess solution was drained with filter paper and the grid was allowed to dry at room temperature. Images were collected using a 4k × 4k Ceta CMOS Camera coupled on a Talos F200C Transmission Electron Microscope (200 kV, Thermo Fisher Scientific) at LNNano/CNPEM. For SEM, *A. castellanii* cultivated in PYG medium were infected with an MOI of 0.01 for 24 hours. The cells were then collected, transferred to a coverslip containing poly-L-lysine, and fixed

with 2.5% glutaraldehyde + cacodylate for 2 hours at room temperature. Subsequently, fixation was performed with 1% osmium tetroxide, washing with 0.1 M cacodylate buffer, and immersion in 0.1% tannic acid. Then, dehydration was performed using serial passages in ethanol solutions with different concentrations. Subsequently, a critical point drying process using CO_2 was carried out. Finally, the samples were accommodated in metal supports (called stubs) and metalized with a layer of gold. Image analyses were performed using a scanning electron microscope (FEI Quanta 200 FEG).

For the one-step-growth curve assay, *A. castellanii* cells were infected in duplicates with an MOI of 10 to obtain a synchronous cycle. After 30 minutes of adsorption, the inoculum was removed, and fresh PYG medium was added. The collections of the supernatant with the cells were performed at 0, 1, 3, 6, 9, 12, 24, 48, and 72 hpi, considering the time 0 hpi right after the adsorption. All times were titrated later, and the curve was constructed from the titration result.

Sequencing, assembly, annotation, and phylogenetic analyses

The samples containing the purified virus were sequenced twice using the equipment Illumina MiSeq, with a paired-end library using the kit Illumina DNA Prep (Illumina Inc., San Diego, CA, USA). First, a *de novo* assembly was performed using the SPAdes 3.13.1 software (36). To increase the *de novo* assembly, the SOAPdenovo2 1.12 (37) program was used. Subsequently, a reference genome (the best hit, Brazilian cedratvirus) was used in the Medusa 1.3 program (38) and the final genome was obtained. For the prediction of the open reading frames (ORFs), the GeneMarkS 4.28 software (39) was used, and the sequences smaller than 50 amino acids were removed from the analyses. Gene prediction using GeneMarkS was performed using both prokaryotic and viral parameters. Since the results were very similar, we opted to use data from prokaryotic parameters because several studies on giant viruses employed this strategy. Therefore, some c. pambiensis (and other pitho-like viruses) genes were predicted to start with alternative start codons, different from ATG. The functional annotation of the predicted proteins was performed using BLASTp against the NCBI NR database considering 1e−5 e-value. The annotation was also done for the six largest clusters of genes using HHpred, with similar results.

For the construction of the phylogenetic trees, the amino acid sequences of the viruses of interest were obtained by the BLASTp tool (default parameters) from the NCBI Genbank database (40). These sequences were aligned with that of c. pambiensis using the MUSCLE 3.8.1551 software (41). The phylogenetic tree was built by the IQtree 1.6.12 program (42) using the best-fitted VT+F+R5 model for amino acid substitution and the likelihood-based method aLRT SH-like with 1,000 pseudoreplicates to estimate branch support values. As aforementioned, substantial variation in coverage was observed in genes belonging to the six largest clusters. This result posed challenges to our phylogenetic analyses due to the potential lack of negative selection in pseudogenes that leads to sequence degeneration. To improve the reliability of our analyses, we defined a cut-off point, in which the gene with the longest CDS was considered as a reference and all the other genes with a CDS shorter than half the size of this gene were removed from the alignment. For cluster 2, the procedure was a little different, as it has three genes much longer than all the others (mean of 427.5% larger than the family median). Therefore, for this cluster, the three largest genes were removed from the analysis and the fourth largest one was considered as a reference.

Detection and mapping of duplicated genes

To detect duplicated genes, we BLASTp the predicted proteins of c. pambiensis against themselves, and hits with coverage ≥30% and e-value $<1e^{-4}$ were considered paralogs. Higher stringent cutoffs were evaluated (i.e., 50%) revealing very similar results. To cluster the paralogs, the Gephi 0.9.7 software (43) was used, based on the list of hits obtained in the previous step. To better understand the duplication events happening within the

genome, we constructed phylogenetic trees for the six largest gene families (more than 20 genes) using amino acid sequences. The programs and parameters used to build these trees were the same used previously.

With the groups of paralogs established, we decided to investigate how these genes spread within the genome. The gene and protein prediction steps gave us the coordinates of each gene, so we developed an R script to construct a Circos plot-like using the circlize 0.4.15 package (44) that draws a line between paralog genes.

Statistical analysis

Spearman correlations were used to assess correlations between genome size and total number of paralogs or total number of genes with a significance level of $P < 0.05$. Data distribution was assessed by the Shapiro-Wilk test. Testing and plotting results were all done in Rstudio (45).

ACKNOWLEDGMENTS

We thank the Laboratório de Vírus of Universidade Federal de Minas Gerais for all the support provided and the Microscopy Center of UFMG, in particular the technicians Denilson Cunha, Rodrigo Ferreira, Altair Mendes, Thalita Arantes, Marilene Oliveira, and Breno Moreira, who collaborated from the preparation to the session for observation of the samples.

We would also like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and Pró-Reitorias de Pesquisa e Pós-Graduação da UFMG (PRPG-UFMG) for the financial support, and LNNano/CNPEM for access to the EM facility via project 20230751. J.S.A., J.P.A.J., and F.R.S. are CNPq researchers.

T.B.M., L.E.D.B., and J.S.A. designed the study and experiments. All authors performed experiments and/or analyses. All authors wrote the manuscript. The text has been entirely written by the authors, and English revision was partially performed using artificial intelligence. All authors approved the final manuscript.

AUTHOR AFFILIATIONS

[1]Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil
[2]Laboratório de Genômica Evolutiva, Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil
[3]Laboratório de Virologia, Departamento de Microbiologia e Imunologia, Instituto de Biotecnologia, Universidade Estadual Paulista (UNESP), Botucatu, Brazil
[4]Del-Bem Lab, Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil
[5]Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Brazil
[6]Centre de Recherche du Centre Hospitalier Universitaire de Québec- Université Laval, Laval, Québec, Canada
[7]Laboratório de Microbiologia Molecular, Universidade Feevale, Novo Hamburgo, Brazil
[8]Archaeal Virology Unit, Institut Pasteur, Université Paris Cité, CNRS UMR6047, Paris, France
[9]Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA
[10]Center for Emerging, Zoonotic, and Arthropod-Borne Infectious Disease Virginia Tech, Blacksburg, Virginia, USA

AUTHOR ORCIDs

Rodrigo A. L. Rodrigues http://orcid.org/0000-0001-7148-4012
Frank O. Aylward http://orcid.org/0000-0002-1279-4050
Luiz-Eduardo Del-Bem http://orcid.org/0000-0001-8472-4476
Jônatas S. Abrahão http://orcid.org/0000-0001-9420-1791

FUNDING

Funder	Grant(s)	Author(s)
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)	303680/2022-9	Jônatas S. Abrahão
Ministério da Ciência, Tecnologia e Inovação (MCTI)	405249/2022-5, 406441/2022-7	Jônatas S. Abrahão
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)	88882.348380/2010-1	Jônatas S. Abrahão

AUTHOR CONTRIBUTIONS

Talita B. Machado, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft | Agnello C. R. Picorelli, Data curation, Formal analysis, Investigation | Bruna L. de Azevedo, Formal analysis, Writing – original draft | Isabella L. M. de Aquino, Formal analysis, Investigation, Writing – original draft | Victória F. Queiroz, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft | Rodrigo A. L. Rodrigues, Formal analysis, Methodology, Writing – original draft | João Pessoa Araújo Jr., Methodology, Writing – original draft | Leila S. Ullmann, Methodology | Thiago M. dos Santos, Methodology, Writing – original draft | Rafael E. Marques, Methodology, Writing – original draft | Samuel L. Guimarães, Methodology, Writing – original draft | Ana Cláudia S. P. Andrade, Investigation, Writing – original draft | Juliana S. Gularte, Methodology, Writing – original draft | Meriane Demoliner, Methodology, Writing – original draft | Micheli Filippi, Methodology, Writing – original draft | Vyctoria M. A. G. Pereira, Investigation, Writing – original draft | Fernando R. Spilki, Methodology, Writing – original draft | Mart Krupovic, Investigation, Methodology, Validation, Writing – original draft | Frank O. Aylward, Investigation, Writing – original draft | Luiz-Eduardo Del-Bem, Conceptualization, Funding acquisition, Investigation, Methodology, Validation, Writing – original draft.

DATA AVAILABILITY

The genome of cedratvirus pambiensis is available at GenBank under accession number OR343515. The genome sequence (fasta) is also available at our research group website (https://5c95043044c49.site123.me/sup-material-of-papers/sup-material-gene-duplication-as-a-major-force-driving-the-genome-expansion-in-some-giant-viruses).

REFERENCES

1. Krebs JE, Goldstein ES, Kilpatrick ST. 2017. Lewin's genes twelve. Jones & Bartlett Learning.
2. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4:10. https://doi.org/10.1186/1471-2229-4-10
3. Reams AB, Neidle EL. 2004. Selection for gene clustering by tandem duplication. Annu Rev Microbiol 58:119–142. https://doi.org/10.1146/annurev.micro.58.030603.123806
4. Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature 449:54–61. https://doi.org/10.1038/nature06107
5. Persi E, Wolf YI, Karamycheva S, Makarova KS, Koonin EV. 2023. Compensatory relationship between low-complexity regions and gene paralogy in the evolution of prokaryotes. Proc Natl Acad Sci U S A 120:e2300154120. https://doi.org/10.1073/pnas.2300154120
6. Li WH, Gu Z, Wang H, Nekrutenko A. 2001. Evolutionary analyses of the human genome. Nature 409:847–849. https://doi.org/10.1038/35057039
7. Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298. https://doi.org/10.1016/S0169-5347(03)00033-8
8. Scola BL, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A giant virus in amoebae. Science 299:2033–2033. https://doi.org/10.1126/science.1081867
9. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: amoeba viruses with Genomes up to 2.5 MB reaching that of parasitic eukaryotes. Science 341:281–286. https://doi.org/10.1126/science.1239181

10. Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. 2018. Tailed giant tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat Commun 9:749. https://doi.org/10.1038/s41467-018-03168-1
11. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. 2020. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat Commun 11:1710. https://doi.org/10.1038/s41467-020-15507-2
12. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D. 2009. Giant marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A 106:21848–21853. https://doi.org/10.1073/pnas.0911354106
13. Legendre M, Fabre E, Poirot O, Jeudy S, Lartigue A, Alempic J-M, Beucher L, Philippe N, Bertaux L, Christo-Foroux E, Labadie K, Couté Y, Abergel C, Claverie J-M. 2018. Diversity and evolution of the emerging pandoraviridae family. Nat Commun 9:2285. https://doi.org/10.1038/s41467-018-04698-4
14. Filée J. 2015. Genomic comparison of closely related giant viruses supports an accordion-like model of evolution. Front Microbiol 6:593. https://doi.org/10.3389/fmicb.2015.00593
15. Filée J, Pouget N, Chandler M. 2008. Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. BMC Evol Biol 8:320. https://doi.org/10.1186/1471-2148-8-320
16. Filée J, Chandler M. 2008. Convergent mechanisms of genome evolution of large and giant DNA viruses. Res Microbiol 159:325–331. https://doi.org/10.1016/j.resmic.2008.04.012
17. Yutin N, Wolf YI, Koonin EV. 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology 466–467:38–52. https://doi.org/10.1016/j.virol.2014.06.032
18. Koonin EV, Yutin N. 2019. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. Adv Virus Res 103:167–202. https://doi.org/10.1016/bs.aivir.2018.09.002
19. Suhre K. 2005. Gene and genome duplication in acanthamoeba polyphaga Mimivirus. J Virol 79:14095–14101. https://doi.org/10.1128/JVI.79.22.14095-14101.2005
20. Moradkhani K, Préhu C, Old J, Henderson S, Balamitsa V, Luo H-Y, Poon M-C, Chui DHK, Wajcman H, Patrinos GP. 2009. Mutations in the paralogous human α-globin genes yielding identical hemoglobin variants. Ann Hematol 88:535–543. https://doi.org/10.1007/s00277-008-0624-3
21. Wickstead B, Gull K. 2011. The evolution of the cytoskeleton. J Cell Biol 194:513–525. https://doi.org/10.1083/jcb.201102065
22. Malik Ghulam M, Catala M, Reulet G, Scott MS, Abou Elela S. 2022. Duplicated ribosomal protein paralogs promote alternative translation and drug resistance. Nat Commun 13:4938. https://doi.org/10.1038/s41467-022-32717-y
23. Koonin EV, Dolja VV, Krupovic M. 2022. The logic of virus evolution. Cell Host Microbe 30:917–929. https://doi.org/10.1016/j.chom.2022.06.008
24. Butkovic A, Dolja VV, Koonin EV, Krupovic M. 2023. Plant virus movement proteins originated from jelly-roll capsid proteins. PLoS Biol 21:e3002157. https://doi.org/10.1371/journal.pbio.3002157
25. Willemsen A, Zwart MP, Higueras P, Sardanyés J, Elena SF. 2016. Predicting the stability of homologous gene duplications in a plant RNA virus. Genome Biol Evol 8:3065–3082. https://doi.org/10.1093/gbe/evw219
26. Chaudhari HV, Inamdar MM, Kondabagil K. 2021. Scaling relation between genome length and particle size of viruses provides insights into viral life history. iScience 24:102452. https://doi.org/10.1016/j.isci.2021.102452
27. Filée J. 2018. Giant viruses and their mobile genetic elements: the molecular symbiosis hypothesis. Curr Opin Virol 33:81–88. https://doi.org/10.1016/j.coviro.2018.07.013
28. Sun C, Feschotte C, Wu Z, Mueller RL. 2015. DNA transposons have colonized the genome of the giant virus pandoravirus salinus. BMC Biol. 13:38. https://doi.org/10.1186/s12915-015-0145-1
29. Krupovic M, Dolja VV, Koonin EV. 2023. The virome of the last eukaryotic common ancestor and eukaryogenesis. Nat Microbiol 8:1008–1017. https://doi.org/10.1038/s41564-023-01378-y
30. Bisio H, Legendre M, Giry C, Philippe N, Alempic J-M, Jeudy S, Abergel C. 2023. Evolution of giant pandoravirus revealed by CRISPR/Cas9. Nat Commun 14:428. https://doi.org/10.1038/s41467-023-36145-4
31. Koonin EV, Krupovic M, Yutin N. 2015. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. Ann N Y Acad Sci 1341:10–24. https://doi.org/10.1111/nyas.12728
32. Rodrigues RAL, Andreani J, Andrade ACDSP, Machado TB, Abdi S, Levasseur A, Abrahão JS, La Scola B. 2018. Morphologic and genomic analyses of new isolates reveal a second lineage of cedratviruses. J Virol 92:e00372-18. https://doi.org/10.1128/JVI.00372-18
33. Koonin EV, Krupovic M. 2018. The depths of virus exaptation. Curr Opin Virol 31:1–8. https://doi.org/10.1016/j.coviro.2018.07.011
34. Machado TB, de Aquino ILM, Abrahão JS. 2022. Isolation of giant viruses of *Acanthamoeba castellanii*. Curr Protoc 2:e455. https://doi.org/10.1002/cpz1.455
35. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. Am J Epidemiol. 27:493–497. https://doi.org/10.1093/oxfordjournals.aje.a118408
36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021
37. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18
38. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDUSa: a multi-draft based scaffolder. Bioinformatics 31:2443–2451. https://doi.org/10.1093/bioinformatics/btv171
39. Besemer J, Borodovsky M. 2005. Genemark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–4. https://doi.org/10.1093/nar/gki487
40. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2014. Genbank. Nucleic Acids Res 42:D32–7. https://doi.org/10.1093/nar/gkt1030
41. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
42. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa131
43. Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. ICWSM 3:361–362. https://doi.org/10.1609/icwsm.v3i1.13937
44. Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. Brors, circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. https://doi.org/10.1093/bioinformatics/btu393
45. Posit team. 2022. Rstudio: integrated development environment for R, Posit software. Available from: http://www.posit.co

4.3 – Artigo 3– The proteomics of the giant cedratvirus particles reveals unique and shared features with pitho-like viruses

Este artigo está em fase final de escrita e será submetido no periódico.Journal of Virology

O cedratvírus é um vírus gigante pertencente à ordem Pimascovirales, um grupo que inclui outros vírus como os pitovírus e os orfeovírus. Os cedratvírus infectam a Acanthamoeba castellanii e possuem um genoma de aproximadamente 600 kb, codificando cerca de 800 proteínas previstas. Neste estudo, realizamos, pela primeira vez, uma análise proteômica de partículas purificadas de cedratvírus, identificando 266 proteínas associadas ao vírus. Essas proteínas estavam envolvidas em uma variedade de processos celulares, incluindo regulação da transdução de sinal, transcrição, processamento de RNA, replicação de DNA, recombinação e reparo, entre outros. Além disso, identificamos 172 proteínas associadas ao hospedeiro nas partículas de cedratvírus. Uma análise comparativa do proteoma do cedratvírus com o do pitovírus sibericum revelou que 89 proteínas são compartilhadas entre os dois vírus, enquanto 96 proteínas são exclusivas do cedratvírus. Essas descobertas fornecem a primeira caracterização proteômica abrangente de um cedratvírus, aprimorando nossa compreensão de sua biologia molecular e destacando características conservadas e exclusivas dentro dos pimascovírus, o que pode orientar estudos futuros sobre a evolução viral e as interações vírus-hospedeiro.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Check for updates

| Editor's Pick | Genomics and Proteomics | Full-Length Text

The proteomics of the giant cedratvirus particles reveals unique and shared features with pitho-like viruses

Talita B. Machado,[1] Talita D. Melo-Hanchuk,[2] Bruna L. de Azevedo,[1] Isabella L. M. de Aquino,[1,3] Juliana R. Cortines,[4] Rafael E. Marques,[2] Jônatas S. Abrahão[1]

AUTHOR AFFILIATIONS See affiliation list on p. 18.

ABSTRACT Cedratvirus is a giant virus belonging to the order *Pimascovirales*, a group that includes other viruses such as pithoviruses and orpheoviruses. Cedratviruses infect *Acanthamoeba castellanii* and have a genome of ~600 kb, encoding ~800 predicted proteins. In this study, we performed, for the first time to our knowledge, a proteomic analysis of purified cedratvirus particles, identifying 266 proteins associated with the virus. These proteins were involved in a variety of cellular processes, including signal transduction regulation, transcription, RNA processing, DNA replication, recombination, and repair, among others. Additionally, we identified in cedratvirus particles 172 host-associated proteins. A comparative analysis of the cedratvirus proteome with that of pithovirus sibericum revealed that 89 proteins are shared between the two viruses, while 96 proteins are unique to cedratvirus. These findings provide the first comprehensive proteomic characterization of a cedratvirus, advancing our understanding of its molecular biology and highlighting both conserved and unique features within the pimascoviruses, which may inform future studies on viral evolution and host-virus interactions.

IMPORTANCE There are still only a few studies on the proteomics of giant viruses. Research published to date has revealed the presence of several proteins involved in processes such as transcription (including RNA polymerase subunits, helicases, and transcription factors), DNA topology regulation (e.g., topoisomerases), and metabolic pathways (e.g., enzymes that help the virus cope with host oxidative stress). Proteomic studies of giant viruses are important, as understanding the role of these proteins can provide information about viral biology, host-virus interactions, and viral evolution. Here, we present, for the first time to our knowledge, a proteomic analysis of a cedratvirus, highlighting unique and shared features with pithovirus, a relative.

KEYWORDS proteomics, giant virus, *Pithoviridae*, cedratvirus pambiensis

The discovery of giant viruses has generated significant interest in the scientific community due to their unprecedented particle size and the complexity of their genomes (mimivirus, Marseillevirus, Pandoravirus, pithovirus, Faustovirus, mollivirus, cedratvirus, kaumoebavirus, Pacmanvirus, Orpheovirus, and Medusavirus) (1–11). In 2014, the isolation of Pithovirus sibericum captured widespread attention, particularly because of its large particle size, measuring approximately 1,300 nm in length (4). Similar to pithovirus, its relatives cedratviruses and orpheoviruses feature large viral particles, ranging in size from 900 to 1,400 nm (7, 10). Four years later, the discovery of Tupanvirus, which has a similarly large particle size ranging from 1,200 to 2,300 µm, further expanded our understanding of these exceptional viruses. In addition to their impressive size, Tupanvirus revealed a sophisticated translational apparatus, underscoring the complexity of giant viruses (12). Since these viruses possess numerous genes that are not

Editor Kristin N. Parent, Michigan State University, East Lansing, Michigan, USA

Address correspondence to Jônatas S. Abrahão, jonatas.abrahao@gmail.com, or Rafael E. Marques, RAFAEL.MARQUES@Inbio.cnpem.br.

Talita B. Machado and Talita D. Melo-Hanchuk contributed equally to this article. Author order was determined alphabetically.

The authors declare no conflict of interest.

Received 9 May 2025
Accepted 14 October 2025
Published 4 December 2025

only novel to the virosphere but also unique in their functions, studying the proteome of giant viruses' particles has become essential to understand their biology.

While proteomic studies on giant viruses remain relatively limited, these investigations are essential for expanding our understanding of the virus-host interactions, as well as the ecological roles these viruses play in nature. For instance, proteomic analyses of giant viruses have revealed the presence of proteins involved in transcription and DNA repair, suggesting that these viruses are capable of initiating genome replication and transcription with a certain degree of autonomy shortly after infecting their host (2, 3, 12–15). This ability to manage essential metabolic processes is a defining characteristic of many giant viruses, setting them apart from smaller, more traditional viruses. A particularly fascinating feature of giant viruses is the high prevalence of ORFans (open reading frames with no detectable homologs), which are genes encoding proteins with unknown functions. For example, in the genome of mimivirus (APMV), ORFans account for approximately 67% of the viral genome (16), while in Pandoravirus salinus, they make up about 93% (3). These ORFans, which are often identified in proteomic analyses, represent a significant mystery in the field. Since their functions remain unknown, it is challenging to predict their roles or importance in the viral life cycle. Proteomic studies are therefore important in confirming the existence of ORFans predicted by bioinformatics, helping to establish their presence in the viral particle and guiding future functional investigations.

In this study, we focus on cedratvirus pambiensis, a giant virus with a genome of 623,564 base pairs and 842 predicted proteins (17). This is the first proteomic analysis of cedratvirus particles according to our knowledge. We found proteins involved in a variety of processes, including signal transduction, transcription and RNA processing, DNA replication, recombination, and repair, as well as various metabolic pathways. Additionally, we performed a comparative analysis with cedratvirus relative, pithovirus sibericum, highlighting the conserved and possibly important roles of certain viral proteins. Taken together, our data contribute to our understanding of the proteomic landscape of cedratvirus pambiensis and its evolutionary relationship with other pitho-like viruses.

RESULTS

Proteomic characterization of edratvirus pambiensis

In this study, we conducted a comprehensive proteomic analysis of purified cedratvirus pambiensis particles, resulting in the identification of 266 proteins (Table 1), which accounts for approximately 31.6% of the proteins predicted by Machado et al. (17). The identified proteins are involved in a wide array of essential cellular processes, highlighting the complexity of this giant virus. Functional characterization of these proteins revealed their involvement in key viral and host-associated mechanisms, including signal transduction regulation, DNA replication, recombination and repair, transcription, and RNA processing (Fig. 1a; Table 1; see Table S1 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-proteomics-of-the-giant-cedratvirus-particles-reveals-unique-and-shared-features-with-pitho-like-viruses). Notably, the presence of proteins associated with DNA repair and replication suggests that the virus may be capable of managing key processes soon after infecting the host cell, similar to other giant viruses. Additionally, the identification of proteins involved in transcription and RNA processing suggests that cedratvirus pambiensis may also regulate its own gene expression, which is consistent with electron microscopy findings indicating that the replication cycle of cedratvirus pambiensis takes place in the amoebal cytoplasm, probably without requiring the amoeba's transcription machinery. Although cedratvirus pambiensis has 76 predicted ORFans (9.03% of its genome), none of them were detected in our proteome analyses.

TABLE 1 Proteins identified in cedratvirus pambiensis particles and their functions inferred from BLASTp best hits

Gene ID	Function	Mol. weight (kDa)	Score	LFQ media	Unique peptides	Sequence coverage (%)
30	Serine/Threonine protein kinase [cedratvirus Ce7-1]	48.493	323.31	23,05845451	8	25.2
50	DNA-directed RNA polymerase subunit RPB2 [Brazilian cedratvirus IHUMI]	136.77	323.31	25,54757373	13	49
51	Endoribonuclease L-PSP/chorismate mutase-like protein [Brazilian cedratvirus IHUMI]	19.794	323.31	28,92235247	4	50.8
169	Uncharacterized protein Ce0701_0016 [cedratvirus Ce7-1]	37.864	323.31	27,48398081	15	45.2
181	DNA-directed RNA polymerase subunit [cedratvirus Ce7-1]	107.79	323.31	26,93042056	29	69
183	Uncharacterized protein Ce0701_0029 [cedratvirus Ce7-1]	31.819	323.31	30,96556981	7	84.1
191	DNA-directed RNA polymerase subunit RPB1 [Brazilian cedratvirus IHUMI]	107.47	323.31	24,62430064	23	60.9
196	Uncharacterized protein Ce0701_0032 [cedratvirus Ce7-1]	29.706	323.31	26,80596161	7	43.8
203	Hypothetical protein BRZCDTV_384 [Brazilian cedratvirus IHUMI]	60.725	323.31	32,34989802	19	68.3
205	Class three lipase [Brazilian cedratvirus IHUMI]	31.591	323.31	26,01503944	7	47.9
239	Class three lipase [Brazilian cedratvirus IHUMI]	31.913	323.31	27,04893557	7	47.3
267	Nucleotidyl transferase [cedratvirus Ce7-1]	49.593	323.31	25,30985133	12	38.2
275	Hypothetical protein BRZCDTV_349 [Brazilian cedratvirus IHUMI]	12.478	323.31	30,38724327	9	74.3
282	Hypothetical protein BRZCDTV_349 [Brazilian cedratvirus IHUMI]	12.31	323.31	29,76369603	11	89.4
284	Hypothetical protein Cplu_202 [cedratvirus plubellavi]	34.845	323.31	29,55129751	11	43.8
297	Hypothetical protein BRZCDTV_334 [Brazilian cedratvirus IHUMI]	40.504	323.31	29,00968424	26	66.1
306	Hypothetical protein BRZCDTV_336 [Brazilian cedratvirus IHUMI]	16.107	323.31	28,51016935	8	91.5
319	Protein containing kinase domain [Brazilian cedratvirus IHUMI]	64.775	323.31	30,79834239	1	56
328	Hypothetical protein BRZCDTV_316 [Brazilian cedratvirus IHUMI]	35.85	323.31	26,55948957	12	59.7
377	Transmembrane domain-containing protein [cedratvirus Ce7-1]	13.202	323.31	27,28659248	3	37.8
385	Hypothetical protein BRZCDTV_286 [Brazilian cedratvirus IHUMI]	17.351	323.31	29,38215192	9	82.3
406	VV D6-like helicase [Brazilian cedratvirus IHUMI]	124.92	323.31	28,32836914	61	59.9
415	Uncharacterized protein Ce0701_0419 [cedratvirus Ce7-1	30.876	323.31	30,36526362	21	82.8
450	Uncharacterized protein Ce0701_0228 [cedratvirus Ce7-1]	37.397	323.31	31,78368187	7	43.8
476	Uncharacterized protein Ce0201_0148 [cedratvirus Ce2-1]	44.949	323.31	24,75685755	8	58.5
482	Hypothetical protein BRZCDTV_230 [Brazilian cedratvirus IHUMI]	36.597	323.31	27,13981883	13	70.1
483	Uncharacterized protein Ce0701_0243 [cedratvirus Ce7-1]	44.913	323.31	26,5209287	13	69
493	DNA-directed RNA polymerase RPB1 (Domain 5) [cedratvirus Ce7-1]	78.429	323.31	26,76166789	18	41.2
523	Uncharacterized protein Ce0701_0188 [cedratvirus Ce7-1]	33.986	323.31	24,6823171	12	67.6
537	Protein containing kinase domain [Brazilian cedratvirus IHUMI]	41.604	323.31	27,89632352	21	52.2
563	Hypothetical protein [cedratvirus kamchatka]	24.295	323.31	32,43375905	22	60.1
597	Hypothetical protein BRZCDTV_159 [Brazilian cedratvirus IHUMI]	13.058	323.31	30,77889061	10	53.8
608	Hypothetical protein BRZCDTV_159 [Brazilian cedratvirus IHUMI]	39.75	323.31	31,35135206	15	66.2
621	Hypothetical protein BRZCDTV_149 [Brazilian cedratvirus IHUMI]	34.348	323.31	32,57397079	27	77.2
636	7-methylguanosine mRNA capping enzyme [Brazilian cedratvirus IHUMI]	111.53	323.31	27,09770902	62	57.6
651	HMG box domain-containing protein [cedratvirus Ce2-1]	25.722	323.31	31,88154093	18	78.4
665	Hypothetical protein BRZCDTV_123 [Brazilian cedratvirus IHUMI]	23.978	323.31	32,87328593	24	88.6
666	VETF-like early transcription factor large subunit [cedratvirus Ce7-1]	138.81	323.31	26,74043338	35	51.4
677	Uncharacterized protein Ce0701_0166 [cedratvirus Ce7-1]	30.554	323.31	27,47323354	19	62.8
678	Hypothetical protein BRZCDTV_113 [Brazilian cedratvirus IHUMI]	25.857	323.31	26,75631332	15	56.3
679	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	34.176	323.31	27,36411985	25	74.5
684	Ankyrin repeat-containing protein [cedratvirus Ce2-1]	51.637	323.31	27,04169782	33	66.6
706	Serine/Threonine protein kinase [Brazilian cedratvirus IHUMI]	53.856	323.31	27,1447614	18	51.6
725	Lipocalin/cytosolic fatty-acid binding domain-containing protein [cedratvirus Ce7-1]	19.047	323.31	29,46610578	3	71.4
778	Uncharacterized protein Ce0701_0127 [cedratvirus Ce7-1]	43.439	323.31	26,03309123	22	57.8
789	Hypothetical protein BRZCDTV_27 [Brazilian cedratvirus IHUMI]	40.755	323.31	28,72144127	15	43.1
807	F-box domain-containing protein [Brazilian cedratvirus IHUMI]	23.293	323.31	28,89256477	9	61.3
243	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	42.678	314.45	26,34749794	12	58

(*Continued on next page*)

TABLE 1 Proteins identified in cedratvirus pambiensis particles and their functions inferred from BLASTp best hits (*Continued*)

Gene ID	Function	Mol. weight (kDa)	Score	LFQ media	Unique peptides	Sequence coverage (%)
572	Hypothetical protein BRZCDTV_169 [Brazilian cedratvirus IHUMI]	54.627	311.51	26,91512489	22	48.3
780	F-box domain-containing protein [cedratvirus Ce2-1]	23.952	305.22	27,82586288	10	61.5
338	Hypothetical protein BRZCDTV_309 [Brazilian cedratvirus IHUMI]	69.598	290.16	28,44070244	2	73.8
514	Uncharacterized protein Ce0701_0188 [cedratvirus Ce7-1]	34.794	286.85	27,34329224	12	65.1
542	Hypothetical protein BRZCDTV_180 [Brazilian cedratvirus IHUMI]	48.826	284.44	26,97695796	8	46.8
582	Hypothetical protein BRZCDTV_169 [Brazilian cedratvirus IHUMI]	55.841	278.72	25,67181778	19	48.8
104	Hypothetical protein BRZCDTV_453 [Brazilian cedratvirus IHUMI]	33.216	277.02	26,83143552	18	52.7
226	Hypothetical protein Cbor_269 [cedratvirus borely]	43.073	271.66	26,43916893	25	68
671	5'–3' exoribonuclease [Brazilian cedratvirus IHUMI]	37.669	265.69	284.935.201	3	44.7
817	F-box domain-containing protein [cedratvirus Ce7-1]	23.31	264.55	26,57676061	6	60.3
255	Uncharacterized protein Ce0701_0065 [cedratvirus Ce7-1]	42.85	261.67	27,12876002	21	60.4
652	VETF-like early transcription factor large subunit [Cedratvirus Ce7-1]	138.89	258.96	25,65801048	33	50.6
252	Uncharacterized protein Ce0701_0057 [cedratvirus Ce7-1]	22.185	253.35	26,81634903	7	49.5
602	Hypothetical protein BRZCDTV_159 [Brazilian cedratvirus IHUMI]	28.896	253.27	30,74014473	9	64.5
363	Transmembrane domain-containing protein [cedratvirus Ce7-1]	29.582	250.16	29,49365107	11	46.1
325	Uncharacterized protein Ce0701_0269 [cedratvirus Ce7-1]	22.066	248.42	30,79998016	11	69.7
168	F-box domain-containing protein [cedratvirus Ce7-1]	26.953	247.75	26,94149272	19	82.7
840	Hypothetical protein Cbor_452 [cedratvirus borely]	19.809	245.33	27,70615768	16	67.1
186	Uncharacterized protein Ce0701_0032 [cedratvirus Ce7-1]	22	244.98	27,12283007	5	48.3
704	Cyclin dependent kinase 2 [Brazilian cedratvirus IHUMI]	52.175	240.5	26,61201286	19	47
331	Hypothetical protein BRZCDTV_316 [Brazilian cedratvirus IHUMI]	36.037	234.09	25,6137619	7	53
784	Hypothetical protein BRZCDTV_32 [Brazilian cedratvirus IHUMI]	10.174	232.94	27,99599457	10	91.4
223	F-box domain-containing protein [cedratvirus Ce7-1]	20.963	228.87	25,25804647	10	49.2
295	Hypothetical protein BRZCDTV_336 [Brazilian cedratvirus IHUMI]	16.158	225.28	28,81258265	6	85.9
556	Hypothetical protein BRZCDTV_176 [Brazilian cedratvirus IHUMI]	11.751	219.53	24,75868479	3	36.6
624	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	29.203	219.07	29,46417109	10	57.8
232	Hypothetical protein BRZCDTV_389 [Brazilian cedratvirus IHUMI]	30.749	218.81	28,4124368	7	33.5
310	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	43.116	216.64	25,19925563	12	49.3
253	F-box domain-containing protein [cedratvirus Ce7-1]	22.506	210.95	26,06808408	7	58.1
407	Uncharacterized protein Ce0701_0429 [cedratvirus Ce7-1]	43.645	210.14	26,90394338	22	66
286	Ankyrin repeat-containing protein [Brazilian cedratvirus IHUMI]	44.885	207.97	28,31149673	12	43.2
381	Uncharacterized protein Ce0701_0312 [cedratvirus Ce7-1]	29.905	205.84	25,87207603	12	58.7
440	RNA polymerase II RPB5 subunit [cedratvirus Ce7-1]	26.077	203.64	26,69197591	9	64
355	Hypothetical protein BRZCDTV_303 [Brazilian cedratvirus IHUMI]	26.963	198.78	25,07421239	11	54.8
382	Uncharacterized protein Ce0701_0312 [cedratvirus Ce7-1]	21.629	183.86	24,96992302	10	61
345	Uncharacterized protein Ce0701_0258 [cedratvirus Ce7-1]	14.46	182.27	27,81278928	8	82.7
296	Protein kinase domain-containing protein [cedratvirus Ce7-1]	33.744	181.96	27,53800964	2	70.9
264	RNAse III putative (double-stranded ribonuclease) [cedratvirus Ce7-1]	38.544	178.22	25,97494634	12	46.2
207	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	40.815	177.55	26,07261276	10	50.3
447	F-box domain-containing protein [cedratvirus Ce7-1]	13.865	177.54	27,19426727	3	50.4
557	RNAse III [cedratvirus Ce7-1]	38.024	174.18	26,75503922	26	77.4
210	Hypothetical protein BRZCDTV_378 [Brazilian cedratvirus IHUMI]	42.644	173.98	25,57869339	14	58
695	Mannosyl phosphorylinositol ceramide synthase [cedratvirus Ce7-1]	14.81	171.25	25,00150935	6	52.8
312	Uncharacterized protein Ce0701_0280 [cedratvirus Ce7-1]	40.974	170.24	24,52449481	20	64.1
714	Hypothetical protein BRZCDTV_78 [Brazilian cedratvirus IHUMI]	31.572	167.13	25,29549726	18	63.1
598	Hypothetical protein BRZCDTV_163 [Brazilian cedratvirus IHUMI]	17.434	167.09	28,77182961	9	63.5
802	Uncharacterized protein Ce0701_0108 [cedratvirus Ce7-1]	37.163	165.97	28,23850123	18	58.6
129	Uncharacterized protein Ce0201_0281 [cedratvirus Ce2-1]	30.261	160.45	26,9009107	5	43.9
59	Ankyrin repeat-containing protein [Brazilian cedratvirus IHUMI]	15.08	160.43	26,05465571	6	77.1
327	Beta_helix domain-containing protein [cedratvirus Ce7-1]	56.542	158.75	27,9752725	3	75

(*Continued on next page*)

TABLE 1 Proteins identified in cedratvirus pambiensis particles and their functions inferred from BLASTp best hits (*Continued*)

Gene ID	Function	Mol. weight (kDa)	Score	LFQ media	Unique peptides	Sequence coverage (%)
176	Hypothetical protein BRZCDTV_403 [Brazilian cedratvirus IHUMI]	25.612	158.39	28,63932228	13	58
594	Hypothetical protein BRZCDTV_163 [Brazilian cedratvirus IHUMI]	17.746	150.15	28,50766881	9	61.6
65	SET domain-containing protein [Brazilian cedratvirus IHUMI]	29.494	149.92	25,87970988	13	61.7
237	Hypothetical protein BRZCDTV_384 [Brazilian cedratvirus IHUMI]	59.32	145.24	30,13671048	17	64.8
417	F-box domain-containing protein [Brazilian cedratvirus IHUMI]	22.042	141.11	28,52620761	5	22.5
285	Uncharacterized protein Ce0701_0431 [cedratvirus Ce7-1]	15.567	135.05	25,78754361	1	15.5
333	DNA-directed RNA Pol II C-term-like phosphatase [cedratvirus Ce7-1]	25.087	130.42	26,96707726	6	58.3
414	TFIIS transcription elongation factor [cedratvirus Ce7-1]	18.955	129.61	26,53813108	10	62.3
810	Uncharacterized protein Ce0201_0219 [cedratvirus Ce2-1]	13.329	129.38	23,81026204	3	47
244	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	43.173	128.47	23,95179558	13	54
157	NTF2-like domain-containing protein [Brazilian cedratvirus IHUMI]	23.923	124.38	31,30645879	12	60.5
527	Uncharacterized protein Ce0701_0191 [cedratvirus Ce7-1]	13.63	123.82	28,49904442	8	82.9
273	Uncharacterized protein Ce0701_0437 [cedratvirus Ce7-1]	12.462	119.91	26,87107086	6	89
242	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	42.767	118.64	25,97226016	11	54
13	Ankyrin repeat-containing protein [Brazilian cedratvirus IHUMI]	12.878	117.37	26,39565659	8	47.7
393	Hypothetical protein BRZCDTV_278 [Brazilian cedratvirus IHUMI]	31.666	114.92	22,81969388	8	35.7
412	Hypothetical protein BRZCDTV_267 [Brazilian cedratvirus IHUMI]	26.784	113.74	26,3551782	11	57.1
799	Hypothetical protein BRZCDTV_18 [Brazilian cedratvirus IHUMI]	9.669	107.32	24,14959208	1	18.9
430	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	17.029	104.14	23,3359019	5	42.6
479	Hypothetical protein BRZCDTV_227 [Brazilian cedratvirus IHUMI]	14.176	102.54	29,49773407	6	31.7
356	Acyl-CoA N-acyltransferase [Brazilian cedratvirus IHUMI]	26.36	100.65	25,96886508	4	42.5
692	WD40 domain-containing protein [cedratvirus Ce7-1]	28.776	99.479	22,86605263	5	20.7
463	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	10.322	98.409	22,71078491	5	55.1
185	Uncharacterized protein Ce0201_0068 [cedratvirus Ce2-1]	27.121	92.622	21,52076149	4	21.5
550	Hypothetical protein BRZCDTV_181 [Brazilian cedratvirus IHUMI]	23.923	92.165	31,86248144	2	33
398	ADPrib_exo_Tox domain-containing protein [cedratvirus Ce2-1]	16.356	91.819	24,87027295	2	27.9
313	Uncharacterized protein Ce0701_0279 [cedratvirus Ce7-1]	28.649	90.883	24,4877739	14	53.8
77	Ankyrin repeat-containing protein [cedratvirus duvanny]	29.309	90.367	20,86135864	5	23.9
781	Methyltransferase [cedratvirus Ce7-1]	36.24	90.157	21,84418805	4	16.2
610	Hypothetical protein BRZCDTV_157 [Brazilian cedratvirus IHUMI]	18.729	89.501	24,2402916	5	47.3
779	Hypothetical protein BRZCDTV_37 [Brazilian cedratvirus IHUMI]	22.162	87.809	24,19650396	5	29.3
357	Acyl-CoA N-acyltransferase [cedratvirus Ce7-1]	28.102	87.483	25,76701291	14	60.3
787	Uncharacterized protein Ce0701_0118 [cedratvirus Ce7-1]	18.544	87.21	28,21001498	7	64
416	F-box domain-containing protein [cedratvirus Ce7-1]	21.82	86.42	27,81792132	15	74.4
788	Uncharacterized protein Ce0701_0117 [cedratvirus Ce7-1]	12.432	84,907	28,89221191	7	59.6
429	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	22.456	84.618	22,08543587	2	11.8
690	Alpha/beta hydrolase [cedratvirus Ce7-1]	34.115	83.334	24,92466863	10	43.8
394	Uncharacterized protein Ce0701_0322 [cedratvirus Ce7-1]	26.309	83.118	23,03069433	4	25.6
83	DNA-directed RNA polymerase subunit RPB2 [Brazilian cedratvirus IHUMI]	136.67	82.202	24,72762553	13	49.1
625	Hypothetical protein BRZCDTV_145 [Brazilian cedratvirus IHUMI]	18.11	81.675	22,88739713	3	24.8
364	Hypothetical protein BQ3484_129 [cedratvirus A11]	99.456	81.437	23,1552302	1	52.9
745	Phosphoglycerate mutase [cedratvirus Ce7-1]	21.179	80.132	22,37655512	1	24.3
271	Transmembrane domain-containing protein [cedratvirus Ce7-1]	36.598	76.555	25,58768972	10	53.4
603	Hypothetical protein BRZCDTV_158 [Brazilian cedratvirus IHUMI]	25.894	75.774	21,68806712	4	23.8
782	Uncharacterized protein Ce0701_0122 [cedratvirus Ce7-1]	71.157	75.352	21,14392726	3	7.8
201	Ubiquitin ligase [Brazilian cedratvirus IHUMI]	15.905	75.323	20,90581449	5	37.2
44	GTP-binding protein [cedratvirus Ce7-1]	13.659	73.639	20,750,199	4	44.2
573	DNA-directed RNA polymerase RPB10 [Brazilian cedratvirus IHUMI]	99.384	72.64	26,60758527	3	63.2
340	DNA-directed RNA Pol II C-term-like phosphatase [cedratvirus Ce7-1]	25.091	72.236	24,99596914	3	50.5
274	Transmembrane domain-containing protein [cedratvirus Ce7-1]	30.199	72.148	23,23042361	5	36.1

(*Continued on next page*)

TABLE 1 Proteins identified in cedratvirus pambiensis particles and their functions inferred from BLASTp best hits (*Continued*)

Gene ID	Function	Mol. weight (kDa)	Score	LFQ media	Unique peptides	Sequence coverage (%)
112	Uncharacterized protein Ce0701_0410 [cedratvirus Ce7-1]	19.185	71.759	21,61158562	3	13.3
657	5'–3' exoribonuclease [Brazilian cedratvirus IHUMI]	39.065	70.757	24,01133664	3	43.3
838	Uncharacterized protein Ce0701_0091 [cedratvirus Ce7-1]	15.514	70.092	21,94168599	2	23.7
326	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	34.225	69.45	24,39121373	15	59.6
798	Uncharacterized protein Ce0701_0110 [cedratvirus Ce7-1]	11.989	69.342	26,71962865	8	58.3
674	Hypothetical protein BRZCDTV_115 [Brazilian cedratvirus IHUMI]	29.008	67.548	20,87562879	4	20.9
828	DNA topoisomerase IIA [Brazilian cedratvirus IHUMI]	54.397	66.562	20,34713554	4	10.4
318	Uncharacterized protein Ce0701_0276 [cedratvirus Ce7-1]	11.718	66.469	25,65533829	1	27.7
21	F-box domain-containing protein [Brazilian cedratvirus IHUMI]	27.165	65.259	25,79525884	13	50.9
804	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	82.143	65.135	22,30910238	2	18.6
235	Lectin domain-containing protein [cedratvirus Ce7-1]	34.936	63.033	23,1832091	2	17
518	Hypothetical protein ZAZAV_583 [cedratvirus Zaza IHUMI]	18.475	62.937	20,13540936	4	26.3
272	EGF-like domain-containing protein [Brazilian cedratvirus IHUMI]	37.01	62.819	21,51404254	4	15.3
418	RecD/TraA family helicase repair protein [cedratvirus Ce7-1]	76.053	62.613	21,0296491	4	7.2
803	Hypothetical protein BRZCDTV_14 [Brazilian cedratvirus IHUMI]	32.944	61.97	25,38799604	10	41.7
68	Translation elongation factor [cedratvirus A11]	46.14	61.297	21,47631836	4	9.7
127	Hypothetical protein BRZCDTV_433 [Brazilian cedratvirus IHUMI]	28.741	60.441	24,44642194	3	35.7
497	Hydrolase-like domain-containing protein [cedratvirus Ce7-1]	22.002	56.843	21,48840332	4	25.5
380	Hypothetical protein BRZCDTV_290 [Brazilian cedratvirus IHUMI]	10.231	56.476	23,51665878	6	64.4
660	Uncharacterized protein Ce0701_0148 [cedratvirus Ce7-1]	32.736	54.854	25,98714892	10	42.2
372	Transmembrane domain-containing protein [cedratvirus Ce7-1]	17.875	53.991	28,22165108	2	28.2
448	RNA polymerase II RPB5 subunit [cedratvirus Ce7-1]	26.183	53.536	26,33921814	7	54.4
211	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	42.027	53.513	31,84666667	11	52.5
365	Divergent major capsid protein [cedratvirus lena]	54.217	52.685	21,47901535	3	7.7
317	Hypothetical protein BRZCDTV_326 [Brazilian cedratvirus IHUMI]	43.549	52.656	25,33771833	15	41
167	AP-endonuclease [Brazilian cedratvirus IHUMI]	50.515	51.646	20,08758036	4	10.3
724	Hypothetical protein BRZCDTV_68 [Brazilian cedratvirus IHUMI]	13.832	50.42	23,81411743	8	69.2
302	Ribonuclease H-like protein [Brazilian cedratvirus IHUMI]	20.247	48.401	21,0942564	4	27.1
472	Glycosyltransferase family 2 [Brazilian cedratvirus IHUMI]	47.334	46.757	20,24222438	3	6.9
404	Hypothetical protein BRZCDTV_272 [Brazilian cedratvirus IHUMI]	16.765	46.496	20,98172283	2	21.4
361	Hypothetical protein BRZCDTV_300 [Brazilian cedratvirus IHUMI]	24.16	45.893	24,88035266	7	50.2
786	PD-(D/E)XK nuclease [cedratvirus Ce7-1]	28.187	45.459	20,65314102	3	13.7
219	F-box domain-containing protein [cedratvirus Ce7-1]	23.775	45.174	20,67233658	4	23.2
601	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	12.499	44.016	26,43851852	3	48.3
373	Hypothetical protein BRZCDTV_297 [Brazilian cedratvirus IHUMI]	72.442	43.954	24,78686841	1	74.2
626	Hypothetical protein Cduv_22 [cedratvirus duvanny]	12.351	43.916	25,79482714	6	47.7
469	Hypothetical protein BRZCDTV_235 [Brazilian cedratvirus IHUMI]	25.515	43.783	20,82987849	3	16.9
442	Hypothetical protein BRZCDTV_246 [Brazilian cedratvirus IHUMI]	50.79	42.721	22,43225733	1	37.9
693	Uncharacterized protein Ce0701_0376 [cedratvirus Ce7-1]	94.825	42.415	24,14972178	2	32.9
360	Acyl-CoA N-acyltransferase [Brazilian cedratvirus IHUMI]	26.287	42.384	22,87440681	4	42.5
615	ATP-dependent DNA ligase [Brazilian cedratvirus IHUMI]	46.883	42.342	23,90722783	12	26.7
238	RRM domain-containing protein [cedratvirus Ce7-1]	25.76	42.282	20,80888875	2	22.2
368	Transmembrane domain-containing protein [cedratvirus Ce7-1]	13.22	41.707	23,66727638	2	37.8
383	Uncharacterized protein Ce0701_0313 [cedratvirus Ce7-1]	31.976	39.561	24,50901413	14	59.4
362	Transmembrane domain-containing protein [cedratvirus Ce7-1]	72.554	38.761	27,86993027	1	12.5
507	Uncharacterized protein Ce0701_0181 [cedratvirus Ce7-1]	18.801	38.409	27,78802427	8	49.7
209	Hypothetical protein BRZCDTV_379 [Brazilian cedratvirus IHUMI]	43.485	38.189	23,75448926	8	40.9
672	5'–3' exoribonuclease [Brazilian cedratvirus IHUMI]	26.132	38.116	28,54252243	10	33.8
193	Hypothetical protein BRZCDTV_394 [Brazilian cedratvirus IHUMI]	31.923	37.478	27,30957921	6	84.1
199	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	20.407	37.379	21,58468564	2	12.4

(*Continued on next page*)

TABLE 1 Proteins identified in cedratvirus pambiensis particles and their functions inferred from BLASTp best hits (*Continued*)

Gene ID	Function	Mol. weight (kDa)	Score	LFQ media	Unique peptides	Sequence coverage (%)
293	DEAD/SNF2 DNA/RNA helicase [Brazilian cedratvirus IHUMI]	34.224	37.147	24,42343521	13	46.1
307	Protein kinase domain-containing protein [cedratvirus Ce7-1]	33.744	36.985	23,85439746	2	70.9
515	Serine/Threonine protein kinase [cedratvirus Ce7-1]	62.353	36.694	22,57287916	12	23.5
276	Transmembrane domain-containing protein [cedratvirus Ce7-1]	23.245	36.604	22,89136124	3	29.2
747	Uncharacterized protein Ce0201_0009 [cedratvirus Ce2-1]	16.237	34.638	26,8409818	6	36
390	Hypothetical protein BRZCDTV_281 [Brazilian cedratvirus IHUMI]	15.077	32.926	20,66976452	2	20.6
84	Endoribonuclease L-PSP/chorismate mutase-like protein [Brazilian cedratvirus IHUMI]	19.649	32.908	26,71621831	2	44.7
294	DEAD/SNF2 DNA/RNA helicase [cedratvirus Ce7-1]	15.146	32.573	23,40559959	2	69.6
720	Uncharacterized protein Ce0701_0333 [cedratvirus Ce7-1]	19.575	32.133	23,0743262	6	40.7
800	Hypothetical protein BRZCDTV_17 [Brazilian cedratvirus IHUMI]	84.726	31.955	24,19575055	3	77.2
353	Helicase nuclease [cedratvirus Ce7-1]	22.229	31.951	26,502,189	7	51.9
304	Uncharacterized protein Ce0701_0276 [cedratvirus Ce7-1]	11.467	31.702	25,24331474	1	20
323	Hypothetical protein BRZCDTV_321 [Brazilian cedratvirus IHUMI]	23.45	30.945	29,03521156	3	16.7
116	Hypothetical protein BRZCDTV_442 [Brazilian cedratvirus IHUMI]	16.342	30.704	20,40,846697	2	15.3
792	Uncharacterized protein Ce0701_0114 [cedratvirus Ce7-1]	18.605	28.684	20,4864502	2	12.7
806	Deoxynucleoside monophosphate kinase [cedratvirus Ce7-1]	24.151	28.052	23,21419716	9	44.7
574	Caspase-like protein [Brazilian cedratvirus IHUMI]	28.759	27.478	24,33054161	4	44.6
311	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	42.502	27.302	24,14866702	9	33
519	Hypothetical protein Cbor_555 [cedratvirus borely]	94.982	27.282	20,94904709	1	12.9
439	F-box domain-containing protein [Brazilian cedratvirus IHUMI]	13.079	26.895	25,70086225	4	41.4
12	Ribonuclease H [Brazilian cedratvirus IHUMI]	18.06	26.379	20,4560585	2	14.3
675	Uncharacterized protein Ce0701_0164 [cedratvirus Ce7-1]	26.381	25.981	20,10263157	2	12.8
389	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	17.166	25.26	23,7800808	10	50
388	Uncharacterized protein Ce0701_0318 [Cedratvirus Ce7-1]	15.703	24.959	23,95928256	7	60.9
100	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	20.128	24.95	25,17103831	5	21.6
359	Acyl-CoA N-acyltransferase [Brazilian cedratvirus IHUMI]	25.96	24.732	23,87978617	9	39.2
278	Uncharacterized protein Ce0701_0433 [cedratvirus Ce7-1]	17.313	24.707	20,99335384	2	13.2
42	Hypothetical protein BRZCDTV_498 [Brazilian cedratvirus IHUMI]	86.448	24.669	21,33561643	2	25
45	Uncharacterized protein Ce0701_0382 [cedratvirus Ce7-1]	14.441	24.341	21,08392715	2	12.9
558	Hypothetical protein BQ3484_556 [cedratvirus A11]	13.707	23.624	21,03420575	1	12.2
552	Hypothetical protein BRZCDTV_180 [Brazilian cedratvirus IHUMI]	48.719	23.502	25,73977661	6	44.5
516	Hypothetical protein Cbor_435 [cedratvirus borely]	49.965	23.145	23,31884511	12	31.2
646	Uncharacterized protein Ce0701_0148 [cedratvirus Ce7-1]	32.881	21.738	25,12048531	7	40.8
841	Hypothetical protein BRZCDTV_2 [Brazilian cedratvirus IHUMI]	13.074	21.032	24,85709953	6	56.9
386	Hypothetical protein BRZCDTV_285 [Brazilian cedratvirus IHUMI]	22.074	20.586	21,79217211	10	58
762	Alpha/beta hydrolase [cedratvirus borely]	27.427	20.278	19,56126785	1	24.6
801	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	15.818	20.179	24,23126666	8	42.2
700	Lipocalin/cytosolic fatty-acid binding domain-containing protein [cedratvirus Ce7-1]	18.121	19.824	27,90670458	1	63.8
670	Cyclin dependent kinase [cedratvirus Ce7-1]	41.715	18.768	22,2694308	2	11.6
375	Divergent Major Capsid Protein [cedratvirus Ce7-1]	53.984	18.642	24,13393211	10	30.3
387	Patatin-like phospholipase [cedratvirus Ce7-1]	30.064	17.495	22,1608874	6	26.6
37	Hypothetical protein BRZCDTV_503 [Brazilian cedratvirus IHUMI]	91.397	16.936	23,85832723	4	56.5
35	Hypothetical protein Ce0701_0388 [cedratvirus Ce7-1]	90.578	16.677	22,11165301	1	43.2
468	Transmembrane domain-containing protein [Brazilian cedratvirus IHUMI]	21.418	16.066	21,76283582	2	20
405	F-box domain-containing protein [Brazilian cedratvirus IHUMI]	18.611	15.657	23,21444575	6	53.8
584	Caspase-like protein [Brazilian cedratvirus IHUMI]	28.745	15.565	22,87681071	3	36.3
539	TFIIB transcription initiation factor [cedratvirus Ce7-1]	21.079	15.515	23,34710248	1	23.7
766	Hypothetical protein BRZCDTV_44 [Brazilian cedratvirus IHUMI]	14.057	14.655	24,76990954	4	47.6

(*Continued on next page*)

TABLE 1 Proteins identified in cedratvirus pambiensis particles and their functions inferred from BLASTp best hits (*Continued*)

Gene ID	Function	Mol. weight (kDa)	Score	LFQ media	Unique peptides	Sequence coverage (%)
756	Alpha/beta hydrolase [Brazilian cedratvirus IHUMI]	30.275	14.565	22,50302696	4	35
528	Protein containing kinase domain [Brazilian cedratvirus IHUMI]	34.999	14.354	21,88711929	7	30.5
204	Hypothetical protein BRZCDTV_383 [Brazilian cedratvirus IHUMI]	25.554	14.292	22,24746513	6	51.6
824	Ankyrin repeat-containing protein [cedratvirus Ce7-1]	19.785	13.893	21,58187358	2	16.1
224	F-box domain-containing protein [cedratvirus Ce7-1]	23.568	13.462	23,2784551	5	24.3
197	Hypothetical protein BRZCDTV_389 [Brazilian cedratvirus IHUMI]	65.323	13.452	27,11423937	4	62.5
723	Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe-like [Brazilian cedratvirus IHUMI]	39.962	13.413	23,38363266	4	11.4
451	Uncharacterized protein Ce0701_0228 [cedratvirus Ce7-1]	16.856	13.042	29,03384336	4	23.9
541	Uncharacterized protein Ce0701_0203 [cedratvirus Ce7-1]	24.039	12.872	28,55572255	2	32.9
280	Uncharacterized protein Ce0701_0437 [cedratvirus Ce7-1]	12.448	12.106	23,07018026	4	82.6
92	Ankyrin repeat-containing protein [Brazilian cedratvirus IHUMI]	15.108	11.379	22,81090864	4	54.2
454	Transmembrane domain-containing protein [cedratvirus Ce7-1]	18.266	11.037	22,38300387	2	16
58	Hypothetical protein BRZCDTV_486 [Brazilian cedratvirus IHUMI]	15.013	10.781	21,90730794	3	31.3
654	PolyA polymerase reg subunit [cedratvirus Ce7-1]	26.708	10.589	21,39780299	5	24.2
371	Hypothetical protein BRZCDTV_290 [Brazilian cedratvirus IHUMI]	91.446	10.514	23,78207397	5	77.2
434	F-box domain-containing protein [Brazilian cedratvirus IHUMI]	25.204	10.324	22,13642438	5	31.1
358	Hypothetical protein BRZCDTV_300 [Brazilian cedratvirus IHUMI]	28.011	10.174	22,83262189	3	29.9
464	Hypothetical protein BRZCDTV_235 [Brazilian cedratvirus IHUMI]	22.828	5.865	20,78565407	3	19.4
305	DEAD/SNF2 DNA/RNA helicase [cedratvirus Ce7-1]	21.879	3.708	22,59363238	2	42.9
708	Hypothetical protein BRZCDTV_84 [Brazilian cedratvirus IHUMI]	17.511	3.072	20,84053421	2	15.6

A significant proportion of these proteins (30%) were classified as hypothetical or uncharacterized, including proteins with transmembrane domains (Fig. 1a; see Table S1 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-proteomics-of-the-giant-cedratvirus-particles-reveals-unique-and-shared-features-with-pitho-like-viruses). The identified proteins also participate in various metabolic pathways, such as lipid, nucleotide, and carbohydrate metabolism, suggesting potential roles in manipulating the host environment to benefit viral propagation. Notably, two major capsid protein (MCP) genes were detected, and both corresponding proteins were identified through proteomic analysis, highlighting their essential roles in virion assembly and structural architecture. Furthermore, 172 proteins associated with the host *Acanthamoeba castellanii* were identified, indicating that host proteins may be incorporated in the viral particles during the late step of the cycle (Table 2). We acknowledge that it is possible that some of the detected proteins may result from residual contamination or incidental

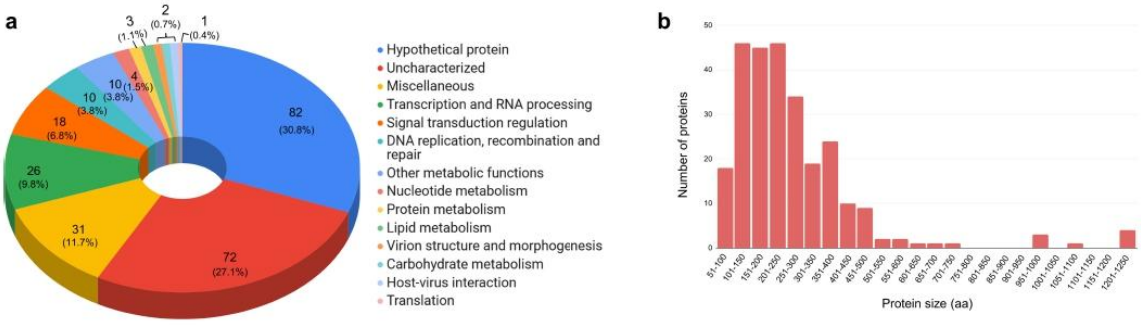

FIG 1 Proteomic characterization of cedratvirus pambiensis. (a) Functional category distribution of proteins identified in the cedratvirus pambiensis proteome. (b) Distribution of proteins according to size (aa).

TABLE 2 *Acanthamoeba* proteins identified in association with cedratvirus pambiensis purified particles after valid value filter and excluding reverse sequences and "only identified by site" entries

Protein IDs	Protein description	Mol. weight (kDa)	Sequence coverage (%)	Unique peptides	LFQ_1	LFQ_2	LFQ_3
gi|470523558|ref|XP_004355037.1|	DIL domain-containing protein	77.039	2.5	1	30,24059	NaN	30,13038
gi|470410604|ref|XP_004336285.1|	Eukaryotic porin protein	33.585	76.5	20	27,36986	27,50718	27,40617
gi|470388901|ref| XP_004334426.1|	Uncharacterized protein ACA1_023990, partial	18.53	47.2	7	26,42153	26,73918	26,54712
gi|470512969|ref|XP_004352564.1|	Actin-1, putative	41.675	65.9	19	26,26202	26,36765	26,56677
gi|302393707|sp|P49633.2|RL40_ACACA	Ubiquitin-ribosomal protein eL40 fusion protein	14.766	33.6	2	26,47314	25,93188	26,39098
gi|470437283|ref|XP_004338972.1|	Adenine nucleotide translocator, putative	34.021	39.3	15	25,87554	26,06163	25,93984
gi|470426286|ref|XP_004337881.1|	Chorismate mutase subfamily protein	20.759	34	4	24,86486	24,71639	25,20083
gi|470527464|ref|XP_004357162.1|	DEAD/DEAH box helicase domain-containing protein, partial	64.87	60.6	34	24,90049	24,93531	24,89458
gi|470482936|ref|XP_004343659.1|	Uncharacterized protein ACA1_099260	44.325	53.5	22	24,75573	24,77047	24,76754
gi|470460317|ref|XP_004341448.1|	Core histone h2a/h2b/h3/h4 superfamily protein	17.322	9.8	2	24,63616	24,93233	24,48401
gi|470408994|ref|XP_004336147.1|	Retrovirus-related Pol polyprotein from transposon TNT 1-94, putative	50.373	2	1	24,65048	24,92452	24,22266
gi|909719706|gb|AKT93988.1|	H(+)-transporting ATPase subunit 9 (mitochondrion)	82.459	12.7	1	24,72931	24,33668	24,5657
gi|470373083|ref|XP_004332988.1|	Peroxidase	36.915	28.3	3	24,40178	24,6019	24,3201
gi|470403734|ref|XP_004335740.1|	Uncharacterized protein ACA1_305680, partial	17,346	7,9	1	24,24663	24,7681	24,29551
gi|470418522|ref|XP_004337098.1|	Cyclophilin, putative	18.1	43.9	7	24,26019	24,55274	24,35763
gi|56405382|sp|Q95VF7.3|PRO1B_ACACA	Acidic profilin IB	12.938	27.8	3	24,45883	24,30536	24,35043
gi|470409867|ref|XP_004336226.1|	Histone H2A variant, putative	14.104	22.9	3	23,88976	23,96517	23,86938
gi|470388239|ref|XP_004334366.1|	C8 sterol isomerase	21.546	21.5	4	23,51629	24,41209	23,78227
gi|470516619|ref|XP_004353105.1|	C2 domain-containing protein	14.171	49.6	5	23,56209	24,12715	23,91752
gi|470484671|ref|XP_004344091.1|	Eukaryotic initiation factor 4A, putative	46.377	57.5	20	23,65086	23,63223	23,68667
gi|470433657|ref|XP_004338597.1|	Peptidase M16 family protein	53.866	48.4	15	23,54215	23,53089	23,61796
gi|470393666|ref|XP_004334908.1|	Cytochrome c oxidase	17.896	26.1	4	23,32924	23,76274	23,3692
gi|6647494|sp|Q37370.1|COX1_ACACA	Cytochrome c oxidase subunit 1+2	99.213	7.8	6	23,42522	23,27938	23,69221
gi|470515801|ref|XP_004352977.1|	Sar1 family small GTPase	21.68	34.2	5	23,31631	23,52325	23,46588
gi|470447037|ref|XP_004340024.1|	Prohibitin 2, putative	32.271	28.9	8	23,44228	23,52468	23,31271
gi|470456089|ref|XP_004341029.1|	Fascin subfamily protein	14.443	41.7	7	23,10238	23,43059	23,45312
gi|470529620|ref|XP_004368102.1|	Prohibitin PHB1, putative	31.311	40.6	9	23,27327	23,39937	23,28673
gi|470401395|ref|XP_004335588.1|	Uncharacterized protein ACA1_359140	65.257	2.2	1	24,05009	22,28452	23,34652
gi|470466481|ref|XP_004341712.1|	Citrate transport family protein	34.287	14.8	4	23,24195	23,14513	23,20725
gi|470426988|ref|XP_004337951.1|	Uncharacterized protein ACA1_297990	15.949	6.1	1	22,5991	23,30966	23,32113
gi|470531418|ref|XP_004358065.1|	Uncharacterized protein ACA1_074190	18.951	16.3	1	23,05793	22,74651	23,37066
gi|470530880|ref|XP_004368325.1|	Peptidase M16 inactive domain-containing protein	63.091	23.2	14	22,92133	23,24747	22,96856
gi|1878344716|gb|QLM05583.1|	Inosineuridine preferring nucleoside hydrolase	40.867	16.8	5	22,93404	23,16197	22,95635
gi|470381459|ref|XP_004333780.1|	EF hand domain-containing protein	24.933	23.3	3	22,94367	22,8694	23,21542
gi|470395572|ref|XP_004335087.1|	Peroxidase	134.07	27.3	26	22,98017	23,05475	22,9362
gi|470426871|ref|XP_004337942.1|	Eukaryotic initiation factor 5a, putative	16.695	19.9	2	22,77211	23,16558	22,92074
gi|470409926|ref|XP_004336232.1|	GTP-binding protein	22.059	5.6	1	22,6285	23,01968	22,81402
gi|470475956|ref|XP_004342208.1|	LIM domain-containing protein	15.264	14.1	1	22,55459	23,10252	22,72228
gi|470510518|ref|XP_004351540.1|	MAPEG domain-containing protein	15.178	18.3	3	22,80069	22,98919	22,56858
gi|470443766|ref|XP_004339681.1|	RNA-binding protein	57.752	20.7	8	22,48375	22,99397	22,78969
gi|470444060|ref|XP_004339708.1|	Rho family, small GTP-binding protein Rac3, putative	21.686	30.1	4	22,60779	22,7918	22,72586
gi|470375092|ref|XP_004333159.1|	High molecular weight heat shock protein	72.287	42.2	20	22,64181	22,85363	22,61131
gi|470414438|ref|XP_004336688.1|	LIM domain-containing protein	18.989	23.7	3	22,24016	23,23182	22,61778
gi|470469092|ref|XP_004341966.1|	Aldehyde dehydrogenase (NAD) family superfamily protein	55.195	18.7	7	22,73705	22,65851	22,66338
gi|470509236|ref|XP_004349728.1|	Rho family, small GTP-binding protein Rac3, putative	21.43	31.6	3	22,78137	22,54948	22,63252
gi|3358312|gb|AAQ16627.1|	Ubiquitin-like protein Ublp94.4	94.409	2.6	1	22,56614	22,73159	22,45861

(*Continued on next page*)

TABLE 2 *Acanthamoeba* proteins identified in association with cedratvirus pambiensis purified particles after valid value filter and excluding reverse sequences and "only identified by site" entries (*Continued*)

Protein IDs	Protein description	Mol. weight (kDa)	Sequence coverage (%)	Unique peptides	LFQ_1	LFQ_2	LFQ_3
gi|470467738|ref|XP_004341828.1|	Elongation factor 1-alpha, putative	41.797	14.9	2	22,54536	22,28949	22,90176
gi|470422038|ref|XP_004337460.1|	Peptidylprolyl cis-trans isomerase	18.521	25.6	4	22,553	22,59099	22,56312
gi|470520651|ref|XP_004353625.1|	Malate dehydrogenase	45.09	44.7	14	22,64926	22,64777	22,40367
gi|470454205|ref|XP_004340817.1|	Uncharacterized protein ACA1_040810	86.587	22.5	2	22,51458	22,58422	22,52834
gi|470373923|ref|XP_004333065.1|	Copper/zinc superoxide dismutase	16.319	55.4	6	22,44995	22,63301	22,41191
gi|470525570|ref|XP_004356622.1|	Mitochondrial phosphate transporter, putative	34.929	22.8	6	22,66303	22,16023	22,66627
gi|470448072|ref|XP_004340125.1|	Leucine rich repeat domain-containing protein	103.73	1.7	2	22,50383	22,38753	22,55852
gi|855099064|gb|AKN79947.1|	Sorting assembly machinery 50 kDa subunit	48.127	21.5	7	22,63969	22,46019	22,30747
gi|470381584|ref|XP_004333791.1|	Gar1 protein RNA-binding region protein	21.336	17.2	3	22,45643	22,51241	22,41158
gi|157833559|pdb|1PRQ|A	Chain A, profilin IA	12.952	20	1	22,44582	22,34557	22,51292
gi|470384632|ref|XP_004334068.1|	Uncharacterized protein ACA1_025730, partial	21.563	32.1	8	22,35714	22,50543	22,43578
gi|470413895|ref|XP_004336630.1|	Syntaxin 6, putative	15.932	11.8	2	22,45588	22,14143	22,69994
gi|470427997|ref|XP_004338048.1|	Peptidase S8 and S53 subtilisin kexin sedolisin, putative	44.359	35	7	22,6054	22,26178	22,2932
gi|470521501|ref|XP_004353726.1|	LIM domain-containing protein	13.774	27.2	2	22,30042	22,47914	22,35512
gi|470458139|ref|XP_004341211.1|	Filamin repeat domain-containing protein	88.769	15.1	10	22,32811	22,34573	22,43311
gi|470492041|ref|XP_004345964.1|	Peptidylprolyl isomerase FKBP12, putative	11.774	33.9	3	NaN	22,3555	22,31797
gi|470429295|ref|XP_004338185.1|	Hypothetical protein ACA1_178210	26.525	36.1	5	22,49782	22,14855	22,31163
gi|470380223|ref|XP_004333668.1|	Ran, putative	24.106	22.5	4	21,89685	22,99229	22,06533
gi|562063|gb|AAD11851.1|	ORF25 (mitochondrion)	14.388	20.2	2	22,17026	NaN	22,396
gi|470423031|ref|XP_004337554.1|	P18, putative	15.161	30.6	5	NaN	22,66173	21,89208
gi|385281908|gb|AFI57874.1|	Beta-tubulin, partial	35.204	9.1	3	NaN	21,85365	22,6846
gi|909719709|gb|AKT93991.1|	H(+)-transporting ATPase subunit 1 (mitochondrion)	57.645	19.2	10	22,18138	22,27247	22,34302
gi|470508831|ref|XP_004349636.1|	Pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase	53.011	8.2	4	22,30516	22,1887	22,29737
gi|470518821|ref|XP_004353386.1|	Uncharacterized protein ACA1_123770	13.5	15.2	1	22,12569	22,15734	22,50082
gi|470430846|ref|XP_004338340.1|	Uncharacterized protein ACA1_203710	22.007	28.7	5	22,27915	22,23216	22,12559
gi|470519565|ref|XP_004353478.1|	rRNA pseudouridine synthase, putative	58.498	10.1	4	22,30941	22,11728	22,20661
gi|470409794|ref|XP_004336219.1|	Protein kinase domain-containing protein	77.397	2	2	22,28102	NaN	22,10623
gi|470375951|ref|XP_004333241.1|	RNA recognition motif domain-containing protein	54.021	24.9	10	21,94628	22,24496	22,26347
gi|470446257|ref|XP_004339937.1|	Exported protein, putative	21.838	15.4	2	22,30791	21,998	22,14622
gi|470377432|ref|XP_004333384.1|	LSM domain-containing protein	14.673	14	2	22,61255	22,17643	21,66049
gi|470530751|ref|XP_004368305.1|	Gamma carbonic anhydrase	29.786	8.1	2	22,09088	22,16809	22,15462
gi|470508654|ref|XP_004349608.1|	Zn-finger in Ran-binding protein and other domain-containing protein	20.926	36.2	4	21,85034	22,43428	22,03564
gi|470449957|ref|XP_004340336.1|	Uncharacterized protein ACA1_371630	24.86	22.4	4	21,86389	22,32056	22,09136
gi|470491417|ref|XP_004345890.1|	PXMP2/4 family protein 3, putative	21.198	18.8	2	22,30316	21,68355	22,25825
gi|470509447|ref|XP_004367727.1|	Autophagy-related protein 27 protein	28.843	37.3	10	21,91984	22,33764	21,88755
gi|440802468|gb|ELR23397.1|	Chorismate mutase, putative	20.869	28.2	4	21,8933	22,0676	22,08352
gi|470506064|ref|XP_004348701.1|	Ras family protein	24.602	21.8	5	21,94806	21,79074	22,23362
gi|470519921|ref|XP_004353526.1|	FGGAP repeat domain-containing protein	48.075	14.9	6	21,96349	21,99824	22,00446
gi|470380103|ref|XP_004333657.1|	Gamma CA3 (gamma carbonic anhydrase 3), putative	30.787	28.4	7	22,07715	22,13389	21,71272
gi|470510220|ref|XP_004351498.1|	I/LWEQ domain-containing protein	70.416	6.5	4	21,94521	21,95326	21,80982
gi|470393565|ref|XP_004334896.1|	Ribosomal protein L6, putative	21.447	22.2	5	21,89149	21,98694	21,81511
gi|881635|gb|AAA93068.1|	Arp3	48.635	19	8	21,8459	22,21715	21,51956
gi|470521487|ref|XP_004353724.1|	Histone H4, putative	12.062	18.2	2	21,75788	22,18435	21,61612
gi|440803274|gb|ELR24182.1|	Cell division control protein 2b, putative	34.498	7.6	2	NaN	21,5582	22,05904
gi|470384600|ref|XP_004334064.1|	Fibrillarin, putative	32.559	21.8	5	21,87763	21,85582	21,68821
gi|470377562|ref|XP_004333398.1|	C2 domain-containing protein	70.216	10.1	6	21,47412	21,98999	21,92644
gi|470521351|ref|XP_004353703.1|	RabE family small GTPase, partial	17.38	21.3	3	21,83881	21,54	21,89586

(*Continued on next page*)

TABLE 2 *Acanthamoeba* proteins identified in association with cedratvirus pambiensis purified particles after valid value filter and excluding reverse sequences and "only identified by site" entries (*Continued*)

Protein IDs	Protein description	Mol. weight (kDa)	Sequence coverage (%)	Unique peptides	LFQ_1	LFQ_2	LFQ_3
gi|470379314|ref|XP_004333569.1|	Uncharacterized protein ACA1_257560	16.537	13.4	2	21,65652	21,84357	21,7615
gi|470413688|ref|XP_004336605.1|	Atp-dependent rna helicase ddx6, putative	46.821	23.8	7	21,74481	21,66917	21,75499
gi|470394048|ref|XP_004334951.1|	Small G-protein	20.63	6	1	21,88372	21,58651	21,46079
gi|470493083|ref|XP_004346472.1|	Cholinesterase	55.495	2.6	1	21,68564	21,44319	21,75316
gi|470449988|ref|XP_004340340.1|	BAR domain-containing protein	46.091	27.1	7	21,66887	21,30697	21,58284
gi|470444553|ref|XP_004339766.1|	Ribosomal protein S8, putative	14.776	20.8	2	21,46154	21,63139	21,45899
gi|470419216|ref|XP_004337172.1|	Mitochondrial import receptor subunit tom40, putative	39.771	11.4	4	21,69098	21,36485	21,48301
gi|470389887|ref|XP_004334520.1|	CMF receptor CMFR1, putative	51.865	20.6	5	21,45332	21,42429	21,62078
gi|470424802|ref|XP_004337738.1|	Casein kinase II subunit alpha, putative	40.052	5.8	2	NaN	21,55112	21,44374
gi|470446444|ref|XP_004339960.1|	SAC3/GANP family protein, partial	72.254	3.4	2	NaN	21,55651	21,43252
gi|470373091|ref|XP_004332989.1|	Peroxidase	23.233	18.1	1	21,46837	21,91903	21,08459
gi|82470775|gb|AAL87229.3|AF480890_1	Metacaspase	50.248	16.9	5	21,41803	21,99948	21,02384
gi|470524449|ref|XP_004356482.1|	Protein phosphatase 2A regulatory B subunit	48.732	9	3	21,39535	21,53365	21,49984
gi|470466515|ref|XP_004341716.1|	Uncharacterized protein ACA1_198450	85.486	15.3	9	21,093	21,48984	21,84142
gi|2169164908|pdb|7RTX|A	Chain A, actophorin	15.422	10.2	1	21,51494	21,0176	21,85829
gi|470420396|ref|XP_004337289.1|	Eukaryotic translation elongation factor 2, putative	93.341	14.8	9	21,25621	21,35875	21,72538
gi|470455700|ref|XP_004340986.1|	Vacuolar proton pump d subunit, putative	35.204	12.2	3	21,29793	21,42019	21,61149
gi|470391452|ref|XP_004334670.1|	ATP:L-methionine S-adenosyltransferase	42.531	12.4	4	NaN	21,63485	21,20467
gi|470526148|ref|XP_004356708.1|	Peroxin 7 (Pex7), putative	36.454	4	1	NaN	21,25442	21,55126
gi|470455432|ref|XP_004340951.1|	Ubiquinol-cytochrome c reductase complex 14 kDa protein	14.346	21.3	2	21,37809	21,67124	21,13959
gi|470415165|ref|XP_004336758.1|	Uncharacterized protein ACA1_390820	33.151	14.6	4	21,37465	21,49116	21,29261
gi|470527031|ref|XP_004367812.1|	1-aminocyclopropane-1-carboxylate deaminase	51.225	4.3	1	21,24928	21,63338	21,25315
gi|470450829|ref|XP_004340430.1|	Elongation factor one alpha, somatic form, putative, partial	35.895	12.8	1	21,26184	21,48104	21,31904
gi|470475309|ref|XP_004342146.1|	ABC2 type transporter superfamily protein	78.702	4.4	3	NaN	21,48055	21,20174
gi|470509344|ref|XP_004350025.1|	Carrier superfamily protein	37.082	14.2	4	21,36068	21,17085	21,43695
gi|470468808|ref|XP_004341938.1|	Domain found in disheveled, egl10, and pleckstrin domain-containing protein	41.043	19.1	6	21,42977	21,23894	21,27407
gi|470423572|ref|XP_004337611.1|	Ras-related protein Rab-2A, putative	22.613	18	3	21,12244	21,22635	21,56118
gi|470444765|ref|XP_004339792.1|	ARP2/3 complex 34 kDa subunit, putative	33.355	21.8	6	20,50368	21,39336	22,01277
gi|470553140|ref|XP_004367444.1|	Alkaline phosphatase family subfamily protein	62.305	2.6	1	20,85673	21,69943	21,30825
gi|470525692|ref|XP_004356641.1|	Protein phosphatase 1, catalytic subunit, alpha, putative	41.035	23.5	6	20,70852	21,66132	21,25828
gi|470401895|ref|XP_004335633.1|	Uncharacterized protein ACA1_037930	21.013	11.3	1	20,77193	21,11579	21,73826
gi|470382483|ref|XP_004333878.1|	Uncharacterized protein ACA1_273770	39.011	25.9	7	20,51384	22,21134	20,80861
gi|470519360|ref|XP_004353460.1|	Uncharacterized protein ACA1_075000	10.377	22	2	21,11858	21,28557	21,07333
gi|470388596|ref|XP_004334397.1|	Protein translation factor, putative	12.289	17.3	2	21,16202	21,29283	20,94714
gi|470510488|ref|XP_004351535.1|	Actin-related protein ARPC3, putative	18.06	14.4	2	21,22199	20,80183	21,3531
gi|470398314|ref|XP_004335307.1|	Sterol 24C-methyltransferase	39.022	15.3	3	21,13295	21,1091	21,11343
gi|470446485|ref|XP_004339965.1|	NADH dehydrogenase, putative	12.6	11.7	1	NaN	21,29227	20,93027
gi|470396670|ref|XP_004335168.1|	BAT1 protein	51.693	14.2	5	20,9047	21,37528	20,99404
gi|470515115|ref|XP_004352880.1|	Uncharacterized protein ACA1_069360	18.977	22.4	3	20,92731	20,73265	21,37311
gi|470421137|ref|XP_004337362.1|	Sec1 family protein, partial	50.699	5.3	2	NaN	21,14283	20,86914
gi|470373753|ref|XP_004333044.1|	Ribosomal protein S5, putative	27.129	25.9	5	20,1706	21,43405	21,34933
gi|470526296|ref|XP_004356732.1|	Vacuolar proton ATPase, putative	92.622	7.4	5	20,90205	20,98197	21,01025
gi|470449103|ref|XP_004340234.1|	Tetratricopeptide repeat domain-containing protein	39.975	5.3	2	20,84276	20,6591	21,37571
gi|470448359|ref|XP_004340152.1|	Carnitine/acylcarnitine translocase, putative	28.804	14.4	3	21,07261	20,7376	20,97967
gi|470455770|ref|XP_004340994.1|	Zn-finger in Ran-binding protein and other domain-containing protein	25.801	21	2	NaN	21,2454	20,58077

(*Continued on next page*)

TABLE 2 *Acanthamoeba* proteins identified in association with cedratvirus pambiensis purified particles after valid value filter and excluding reverse sequences and "only identified by site" entries (*Continued*)

Protein IDs	Protein description	Mol. weight (kDa)	Sequence coverage (%)	Unique peptides	LFQ_1	LFQ_2	LFQ_3
gi|470428729|ref|XP_004338121.1|	Uncharacterized protein ACA1_025540	44.626	9.1	3	NaN	20,71129	21,00319
gi|470425113|ref|XP_004337775.1|	Uncharacterized protein ACA1_379440	11.043	26.3	2	20,98537	21,1354	20,39665
gi|470399342|ref|XP_004335405.1|	Ras subfamily protein	28.833	16.3	2	21,00265	20,4997	20,96148
gi|470467441|ref|XP_004341795.1|	Guanine nucleotide-binding protein beta subunit, putative	35.215	11.8	3	21,20437	20,73397	20,50038
gi|470465131|ref|XP_004341583.1|	Ribosomal protein S9, putative	16.561	8	1	20,59295	20,78596	21,00086
gi|470407043|ref|XP_004336018.1|	Vacuolar protein sorting-associated protein 45, putative	63.206	6.8	4	20,76085	20,53032	20,8753
gi|470413626|ref|XP_004336599.1|	Uncharacterized protein ACA1_321030	56.146	14.2	5	20,39509	21,20305	20,49999
gi|470523597|ref|XP_004355043.1|	Ras-related protein Rab-7, putative	20.543	12.4	2	20,78101	20,56932	20,73942
gi|470427289|ref|XP_004337981.1|	Catalase, putative	55.346	3.5	1	20,77411	20,8285	20,46862
gi|470530269|ref|XP_004368214.1|	Small nuclear ribonucleoprotein Sm D1, putative	13.41	10.4	1	20,85148	20,76709	20,351
gi|470389022|ref|XP_004334436.1|	LIM domain-containing protein	39.651	8.5	2	20,50116	20,31296	21,07679
gi|470484138|ref|XP_004344039.1|	Peroxin 5 (Pex5), putative	87.461	4.5	3	20,54633	20,71087	20,59067
gi|470443710|ref|XP_004339676.1|	Elongation factor SIII p15 subunit, putative	11.513	38.8	4	20,29345	20,75654	20,50814
gi|470532050|ref|XP_004358287.1|	Eukaryotic ribosomal protein L18, putative	20.902	12.9	2	20,70455	20,00031	20,79257
gi|470446370|ref|XP_004339951.1|	Uncharacterized protein ACA1_208120	36.026	4.3	1	20,58536	19,99507	20,89467
gi|470468369|ref|XP_004341892.1|	DNA repair protein RAD51, putative	37.261	21.6	5	20,26819	NaN	20,701
gi|470395029|ref|XP_004335030.1|	Ribosomal family s4e, putative	27.273	9.1	2	NaN	20,60729	20,35348
gi|470391709|ref|XP_004334702.1|	Hras1 protein	22.754	5.4	1	NaN	20,50485	20,37581
gi|470431682|ref|XP_004338423.1|	Calponin domain-containing protein	19.77	13.4	2	20,20347	20,38574	20,54161
gi|470417075|ref|XP_004336955.1|	Translation elongation factor Tu, putative	49.291	11.4	5	19,84011	21,41761	19,74516
gi|470522464|ref|XP_004354032.1|	Succinate dehydrogenase and fumarate reductase ironsulfur protein	32.747	4.1	1	20,46832	20,19508	20,32476
gi|470380149|ref|XP_004333661.1|	Actin bundling protein	31.046	12.6	3	NaN	20,3005	20,35434
gi|440790682|gb|ELR11962.1|	Extracellular response kinase, putative	47.967	9.6	4	19,95035	20,0654	20,56151
gi|470486273|ref|XP_004344589.1|	Ras-like protein 1, putative	22.559	12.7	2	NaN	20,19135	20,12326
gi|470380314|ref|XP_004333676.1|	Proteasome subunit alpha type 6, putative	27.045	14.9	3	NaN	20,2876	19,99755
gi|470475061|ref|XP_004342123.1|	GTP-binding protein	24.385	16.3	3	20,20431	20,17913	19,89812
gi|470449066|ref|XP_004340229.1|	Ribosomal protein L6e, putative	27.47	6.1	1	20,014	19,88247	20,32729
gi|470457495|ref|XP_004341141.1|	Calcium-binding mitochondrial carrier protein	80.548	3.4	2	NaN	19,93603	20,19352
gi|470501765|ref|XP_004347187.1|	Microtubule-associated protein (MAP65/ASE1 family)	61.808	2	1	19,9894	19,79853	19,63835
gi|470422827|ref|XP_004337531.1|	Uncharacterized protein ACA1_163460	24.655	10.6	2	NaN	19,79066	19,70578
gi|470510548|ref|XP_004351544.1|	Phosphatidate cytidylyltransferase	39.059	5.1	1	NaN	19,53448	19,76374
gi|470374283|ref|XP_004333084.1|	AcylCoA dehydrogenase, putative	58.05	6.1	3	NaN	19,58344	19,68368
gi|470441553|ref|XP_004339426.1|	Ribosomal protein L10, putative	23.766	10.1	2	19,31586	NaN	19,6768
gi|470482928|ref|XP_004343658.1|	Uncharacterized protein ACA1_099250	33.006	6.2	2	18,84607	18,81541	19,16226

association with the virion. However, we performed two rounds of sucrose cushion purification to minimize such contaminants, and we carefully interpret our proteomic data with this consideration in mind.

Size variation and genomic distribution of proteins in cedratvirus pambiensis

The protein sizes identified in this study ranged from 52 to 1,225 amino acids, corresponding to 6 to 135 kDa in mass, with the majority of proteins falling within the 101–300 amino acid range (11–33 kDa). Interestingly, no proteins were identified in the particles within the amino acid positions of 751–950, 1001–1050, and 1101–1200, which may reflect a limitation of the technique, as larger proteins have more difficulty entering SDS-PAGE, or it may be the case that certain protein sizes may be underrepresented or absent in the viral proteome (Fig. 1b). Further analysis revealed that most proteins found in the particles were associated with gene IDs in the 301–400 range (53 genes), followed

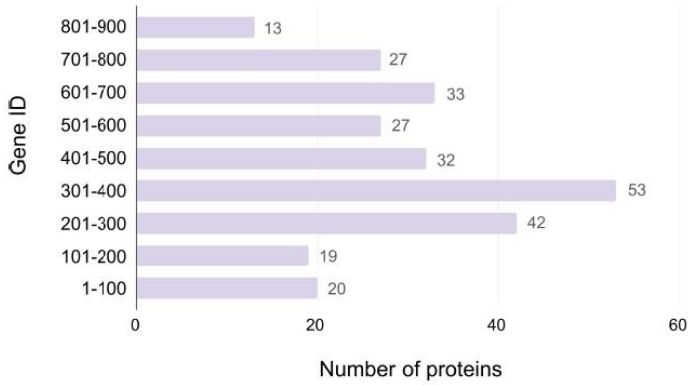

FIG 2 Distribution of the number of proteins in each gene ID range.

by the 201–300 range (42 genes). In contrast, gene IDs in the 901–1000 range exhibited a lower number of identified proteins (Fig. 2).

In addition to protein size and distribution, we analyzed the particle proteome in terms of the orientation of genes on the positive and negative strands of the genome. The results showed a slightly higher number of proteins encoded by the positive strand (n = 144) compared to the negative strand (n = 122). When examining the distribution of proteins across different genomic intervals, we found that the distribution was largely similar between the two strands. However, in the 1–100 kb region, the positive strand encoded a higher number of proteins when compared to the same region in the negative strand. The 201–300 kb region is responsible for encoding the largest number of proteins in the particle, suggesting that this region may be a hotspot for transcription or translation activity during the late times of the cycle (Fig. 3). This uneven genomic distribution of virion-encoded genes has been previously observed in proteomic analyses of Pandoravirus, and our findings for cedratvirus are consistent with this trend.

Comparative proteomic analysis between cedratvirus pambiensis and pithovirus sibericum

Among the viruses clustered with cedratvirus pambiensis, only pithovirus sibericum has published proteomic data, identifying 159 proteins (4). Comparative analysis of the two proteomes revealed 89 shared proteins, predominantly uncharacterized (49 proteins, hypothetical proteins included). Other shared proteins were involved in

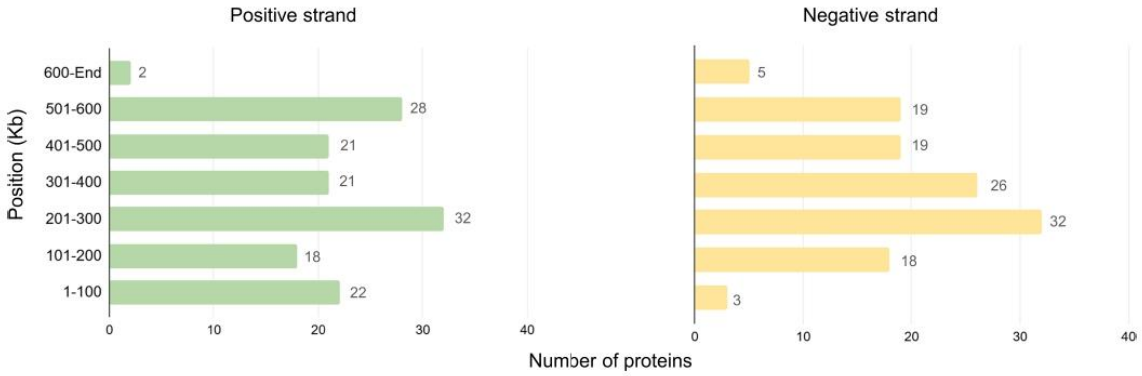

FIG 3 Distribution of genes within the genome, according to their orientation on the positive or negative strands.

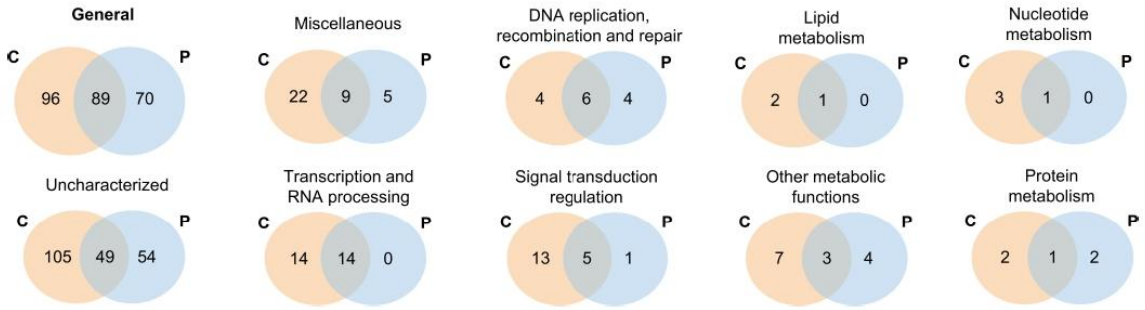

FIG 4 Comparison between cedratvirus pambiensis and pithovirus sibericum proteomes. Eight proteins were not represented because they are only present in the cedratvirus proteome. C: cedratvirus pambiensis. P: pithovirus sibericum.

transcription and RNA processing (14 proteins) and DNA replication, recombination, and repair (6 proteins). Additional shared functional categories included signal transduction regulation (5 proteins), miscellaneous functions (9 proteins), lipid metabolism (1 protein), and other metabolic processes (3 proteins) (Fig. 4).

Pithoviruses contain a single protein related to lipid metabolism and another related to nucleotide metabolism, which are shared with cedratvirus. All of the proteins involved in transcription and RNA processing identified in the pithovirus proteome are shared with cedratvirus pambiensis (Fig. 4). Notably, cedratvirus pambiensis possesses unique proteins involved in carbohydrate metabolism (2), host-virus interactions (2), virion structure and morphogenesis (2), and translation (1), which were not identified in the pithovirus proteome.

We performed BLASTp of all cedratvirus pambienis proteins against pithovirus sibericum proteins and observed that 96 (36.1%) did not have any significant similarity, that is, they are exclusive to cedratvitus (Table S2).

We selected the 20 most abundant proteins in the cedratvirus pambiensis proteome, of which 13 are hypothetical proteins, 4 are uncharacterized, 1 HMG box domain-containing protein, 1 NTF2-like domain-containing protein, and 1 protein containing a kinase domain. BLAST analysis of these proteins against the pithovirus sibericum proteome showed no significant similarity for 6 of them (Table 3). The same procedure applied to the 20 most abundant proteins in the pithovirus proteome, which include 17 hypothetical proteins, 1 DNA-binding ferritin-like protein, 1 DNA-directed RNA polymerase RPB10, and 1 transcription elongation factor TFIIS, revealed that 5 proteins had no significant similarity in the cedratvirus proteome. Additionally, for the proteins pv_6, pv_411, pv_114, and pv_456, more than one corresponding match was identified in the cedratvirus genome (Table 3).

These data demonstrate that the proteomes of cedratvirus pambiensis and ithovirus sibericum have significant differences between them. However, key features are conserved, including a nearly complete set of genes involved in transcription.

DISCUSSION

Since the discovery of APMV in 2003 (1), giant viruses have been isolated worldwide. They stand out due to their large particle sizes and the complexity of their genomes (3, 4, 7, 12, 16, 18). Despite the growing interest in these viruses, many aspects of their life cycle and interaction with the host or the environment still remain a mystery. Proteomic studies of particles published to date have identified the presence of essential proteins that contribute to the partial independence of giant viruses from their hosts, such as proteins involved in transcription and redox control (2, 3, 12–15). One intriguing feature of giant viruses is the presence of ORFans in their genomes, which are genes with unknown functions but are often identified in proteomic analysis. Proteomic studies may

TABLE 3 Comparison between the most abundant proteins in the cedratvirus pambiensis and pithovirus sibericum proteomes and their orthologs

Position	Cedratvirus protein ID	Ortholog in pithovirus ID	Position	Pithovirus protein ID	Ortholog in cedratvirus ID
1	gene_665 [hypothetical protein]	pv_449 [hypothetical protein]	1	pv_449 [hypothetical protein]	gene_665 [hypothetical protein]
2	gene_621 [hypothetical protein]	pv_46 [hypothetical protein]	2	pv_461 [hypothetical protein]	gene_621 [hypothetical protein]
3	gene_563 [hypothetical protein]	pv_79 [hypothetical protein]	3	pv_93 [hypothetical protein]	No significant similarity found
4	gene_203 [hypothetical protein]	pv_106 [hypothetical protein]	4	pv_106 [hypothetical protein]	gene_203 [hypothetical protein]
5	gene_651 [HMG box domain-containing protein]	pv_449 [hypothetical protein]	5	pv_365 [DNA-binding ferritin-like protein]	No significant similarity found
6	gene_550 [hypothetical protein]	pv_423 [hypothetical protein]	6	pv_6 [hypothetical protein]	gene_608; gene_ 602 [hypothetical protein]
7	gene_450 [uncharacterized protein]	pv_45 [hypothetical protein]	7	pv_100 [hypothetical protein]	No significant similarity found
8	gene_608 [hypothetical protein]	pv_6 [hypothetical protein]	8	pv_384 [hypothetical protein]	gene_157 [NTF2-like domain-containing protein]
9	gene_157 [NTF2-like domain-containing protein]	pv_384 [hypothetical protein]	9	pv_45 [hypothetical protein]	gene_450 [uncharacterized protein]
10	gene_183 [uncharacterized protein]	pv_383 [hypothetical protein]	10	pv_265 [hypothetical protein]	gene_176 [hypothetical protein]
11	gene_325 [uncharacterized protein]	No significant similarity found	11	pv_284 [hypothetical protein]	gene_798 [uncharacterized protein]
12	gene_319 [protein containing kinase domain]	pv_418 [protein kinase]	12	pv_160 [hypothetical protein]	No significant similarity found
13	gene_597 [hypothetical protein]	No significant similarity found	13	pv_1 [hypothetical protein]	No significant similarity found
14	gene_602 [hypothetical protein]	pv_6 [hypothetical protein]	14	pv_256 [hypothetical protein]	gene_176 [hypothetical protein]
15	gene_275 [hypothetical protein]	No significant similarity found	15	pv_383 [hypothetical protein]	gene_183 [uncharacterized protein]
16	gene_415 [uncharacterized protein]	No significant similarity found	16	pv_411 [hypothetical protein]	gene_295; gene_306 [hypothetical protein]
17	gene_237 [hypothetical protein]	pv_106 [hypothetical protein]	17	pv_448 [hypothetical protein]	gene_345 [uncharacterized protein]
18	gene_282 [hypothetical protein]	No significant similarity found	18	pv_114 [DNA-directed RNA polymerase RPB10]	gene_573 [DNA-directed RNA polymerase RPB10]; gene_810 [uncharacterized protein]
19	gene_284 [hypothetical protein]	pv_92 [hypothetical protein]	19	pv_29 [transcription elongation factor TFIIS]	gene_414 [TFIIS transcription elongation factor]
20	gene_479 [hypothetical protein]	No significant similarity found	20	pv_456 [hypothetical protein]	gene_363; gene_372; gene_624 [transmembrane domain-containing protein]

help to elucidate the roles and importance of these proteins in viral biology, manipulation of the host environment, and viral evolution.

The proteomic analysis of cedratvirus pambiensis identified 266 viral proteins and 172 proteins associated with the host *Acanthamoeba castellanii*. The presence of proteins involved in transcription and RNA processing suggests that these viruses may be capable of initiating transcription with a certain degree of autonomy soon after infecting their host. Various metabolism-associated proteins were identified, such as those related to lipid, nucleotide, and carbohydrate metabolism, suggesting potential manipulation of the host environment to benefit viral propagation. These findings suggest that cedratvirus pambiensis virions possess a sophisticated set of proteins that likely play a critical role in initiating and maintaining the viral life cycle within the host amoeba.

Structural proteins were also identified, and two MCP genes were detected, highlighting their essential roles in virion assembly. However, unlike what is described for other nucleocytoviruses as mimivirus and phycodnaviruses, the cedratvirus MCP is not among the most abundant proteins found in the capsid. This suggests that cedratvirus MCP proteins may not serve the same role as they do in other nucleocytoviruses. This is compatible with the cedratvirus non-ichosaedrical, oval-shaped particle. In mollivirus, for instance, MCP remains associated with the particle and acts as a scaffold for virion formation, supporting the idea that MCP functions may have diverged among different lineages (19). Furthermore, 172 host-associated proteins were identified, indicating that host proteins may be incorporated into the viral particles during the late step of the cycle.

Comparative analysis of cedratvirus pambiensis and its relative pithovirus sibericum proteomes revealed 89 shared proteins. Most of them are uncharacterized proteins (49 proteins) but were also identified proteins involved in transcription and RNA processing (14 proteins), DNA replication, recombination, and repair (6 proteins), signal transduction regulation (5 proteins), miscellaneous functions (9 proteins), lipid metabolism (1 protein), nucleotide metabolism (1 protein), and other metabolic processes (3 proteins). We also observed 96 (36.1%) proteins that are exclusive to cedratvirus pambiensis proteome, among them proteins involved in carbohydrate metabolism (2), host-virus interactions (2), virion structure and morphogenesis (2), and translation (1). These results represent one more step forward toward the understanding of the core proteome of the oval-shaped pimascoviruses. It seems that pithoviruses and cedratviruses have their cycle entirely in the host cytoplasm, which is compatible with proteomics and electron microscopy findings. The proteomic analysis of Orpheovirus particles will be useful to a better understanding of the core and exclusive proteomics of this group of viruses.

In conclusion, our data demonstrate that cedratvirus pambiensis particles possess a distinct set of proteins, with significant differences compared to the proteome of its relative, ithovirus sibericum. However, key features are conserved, including a nearly complete set of genes involved in transcription, suggesting that this virus can initiate and complete its life cycle without an apparent nuclear phase. Further research on other pimascoviruses is needed to gain a deeper understanding of the essential core set of proteins shared by this virus group.

MATERIALS AND METHODS

Production and purification

Acanthamoeba castellanii cells were infected with cedratvirus pambiensis particles at an MOI of 0.01 in glass culture flasks (300 cm^2) with 35 mL of PYG medium and kept at 30°C in a rotary cell oven for 5 days. After complete cell lysis, the contents of the flasks were collected and centrifuged at 145 × *g* for 10 min. After centrifugation, the supernatant was transferred to a sterile beaker, and the pellet was discarded. This supernatant was added to a sucrose cushion (40%) and then centrifuged at 33,000 × *g* for 1 h, between 4°C and 8°C. The final pellet was resuspended in 10 mL of phosphate-buffered saline, and this content was again added to a sucrose cushion (40%) and centrifuged at 33,000 × *g* for 1 h, between 4°C and 8°C. The centrifugation process was performed twice to obtain the purest sample possible.

Protein preparation

For protein extraction, 100 million viral particles were resuspended in Laemmli buffer (90 mM dithiothreitol [DTT], 2% SDS, 80 mM Tris-HCl pH 6.8, 10% glycerol, and 0.1% bromophenol blue) for protein extraction and solubilization. Proteins were then separated using SDS-PAGE, and electrophoresis was halted as soon as the samples entered the resolving gel. The gel was stained with Coomassie Brilliant Blue (Bio-Rad), and the single band containing all proteins was excised for in-gel digestion. The experiment was performed in triplicate (see Fig. S1 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-proteomics-of-the-giant-cedratvirus-particles-reveals-unique-and-shared-features-with-pitho-like-viruses). The excised gel band was destained using a solution of 50% methanol and 2.5% acetic acid, followed by protein reduction with 10 mM DTT and alkylation with 50 mM iodoacetamide in the dark for 30 min. The reaction was quenched by adding 10 mM DTT and incubating in the dark for 15 min. Protein digestion was performed using 1 µg of trypsin (Promega, Madison, WI, USA) at 37°C for 16 h. The reaction was stopped by adding 5% formic acid, and the peptides were extracted with 5% formic acid in 50% acetonitrile. Peptides were desalted using the StageTips method with C18 Empore disks (3M, St. Paul, MN, USA). After desalting, samples were dried in a vacuum concentrator and reconstituted in 10 µL of 0.1% formic acid for further analysis. The data presented represent technical replicates. The same viral preparation was aliquoted, lysed independently, separated by SDS-PAGE, and processed separately for mass spectrometry analysis.

LC/MS-MS

The samples were analyzed by LC-MS/MS on an Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific, USA) connected to the EASY-nLC 1200 system (Proxeon Biosystems, USA) through a Proxeon nanoelectrospray ion source. Peptides were separated by a 2%–40% acetonitrile gradient over 155 min, using 80% acetonitrile in 0.1% formic acid, in a trap Acclaim PepMap 100 nanoViper 2PK C18 (2 cm × ID75 µm, 3 µm particle size, Thermo Scientific) in line with an analytical PepMap RSLC C18 ES 902 column (50 cm × ID75 µm, 2 µm particle size) at a flow rate of 250 nL/min. The nanoelectrospray voltage was set to 1.7 kV, and the source temperature was 275°C. The full-scan MS spectra (m/z 375–1,500) were acquired in the Orbitrap analyzer after accumulation of a target value of 3e6. Resolution in the Orbitrap was set to r = 60,000, and the 20 most intense peptide ions with charge states ≥2 were sequentially isolated to a target value of 3e5 and fragmented by high-energy collisional dissociation (normalized collision energy of 27%). The signal threshold for triggering an MS/MS event was set to 1e4. Dynamic exclusion was enabled with an exclusion duration of 20 s and a repeat count of 1. The maximum injection time was 60 ms.

Data analysis

Proteins were identified using MaxQuant version 2.4.7.0 against concatenated databases from *Acanthamoeba castellani* and genome sequences (37,160 protein sequences, 15,296,385 residues, April 2024) using the Andromeda search engine. Six-frame translation was performed using the six-frame translation tool embedded in MaxQuant. Carbamidomethylation was set as a fixed modification, and N-terminal acetylation and oxidation of methionine were used as variable modifications. A maximum of two trypsin/P missed cleavages, a tolerance of 10 ppm for precursor mass, and a tolerance of 0.02 Da for fragment ions were set for peptide identification. Protein groups were automatically inferred by the Andromeda engine using the parsimony principle. A maximum of 1% FDR, calculated using reverse sequences, was set for both protein and peptide identification. Protein quantification was performed using the LFQ algorithm implemented in MaxQuant software to reflect a normalized protein quantity deduced from razor + unique peptide intensity values. A minimal ratio count

of one was set. Protein identifications assigned as "Reverse," only identified by "site" and those identified in only one of the samples were excluded from further analysis (see Fig. S2 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-proteomics-of-the-giant-cedratvirus-particles-reveals-unique-and-shared-features-with-pitho-like-viruses). LFQ intensities were log2-transformed in Perseus version 2.0.11 and used in subsequent analyses. Proteins with one or two peptides and coverage <10%, simultaneously, were removed from the analysis.

Bioinformatic analyses

The functional annotation of cedratvirus pambiensis proteins was performed using BLASTp against the NCBI NR database under default parameters. The functional categories utilized in this study for both viruses were based on those described by Yutin et al. (2). Comparative analysis of the viral proteomes was performed using BLASTp, aligning all cedratvirus pambiensis proteins against the published pithovirus sibericum proteome (4), with an e-value threshold of 0.001. For the comparative analysis, the most recent functional description of each protein was considered. The same analysis was conducted for the 20 most abundant pithovirus sibericum proteins, aligning them against all cedratvirus pambiensis proteins. All graphs were generated using Microsoft Excel and PowerPoint software.

ACKNOWLEDGMENTS

The authors thank the Laboratório de Vírus of the Universidade Federal de Minas Gerais for all the support provided.

The authors also thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and Pró-Reitorias de Pesquisa e Pós-Graduação da UFMG (PRPG-UFMG) for the financial support. J.S.A. and R.E.M. are CNPq researchers.

The manuscript was linguistically reviewed with the assistance of artificial intelligence (ChatGPT). This research used facilities of the Brazilian Biosciences National Laboratory (LNBio), part of the Brazilian Center for Research in Energy and Materials (CNPEM), a private non-profit organization under the supervision of the Brazilian Ministry for Science, Technology, and Innovations (MCTI). Adriana Franco Paes Leme and Bianca Alver Pauletti are acknowledged for their assistance during the experiments at the Mass Spectrometry Laboratory Facility (proposal number 20233792).

AUTHOR AFFILIATIONS

[1]Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil
[2]Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Brazil
[3]Centro de Tecnologia de Vacinas (CT Vacinas)/BH-Tec, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil
[4]Departamento de Virologia, Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

AUTHOR ORCIDs

Juliana R. Cortines http://orcid.org/0000-0003-2481-9267
Rafael E. Marques http://orcid.org/0000-0002-6949-0947
Jônatas S. Abrahão http://orcid.org/0000-0001-9420-1791

AUTHOR CONTRIBUTIONS

Talita B. Machado, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Talita D. Melo-Hanchuk, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing – original draft, Writing – review and editing | Bruna L. de Azevedo, Investigation, Methodology | Isabella L. M. de Aquino, Investigation, Methodology | Juliana R. Cortines, Data curation, Writing – review and editing | Rafael E. Marques, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review and editing | Jônatas S. Abrahão, Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The proteomics data have been deposited to the ProteomeXchange database under the accession number PXD069162.

REFERENCES

1. Scola BL, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A giant virus in amoebae. Science 299:2033–2033. https://doi.org/10.1126/science.1081867
2. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D. 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci USA 106:21848–21853. https://doi.org/10.1073/pnas.0911354106
3. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341:281–286. https://doi.org/10.1126/science.1239181
4. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie J-M. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc Natl Acad Sci USA 111:4274–4279. https://doi.org/10.1073/pnas.1320670111
5. Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, Rossmann M, Colson P, Raoult D, La Scola B. 2015. Faustovirus, an Asfarvirus-related new lineage of giant viruses infecting amoebae. J Virol 89:6585–6594. https://doi.org/10.1128/JVI.00115-15
6. Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic J-M, Ramus C, Bruley C, Labadie K, Shmakova L, Rivkina E, Couté Y, Abergel C, Claverie J-M. 2015. In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*. Proc Natl Acad Sci U S A 112:E5327–E5335. https://doi.org/10.1073/pnas.1510795112
7. Andreani J, Aherfi S, Bou Khalil JY, Di Pinto F, Bitam I, Raoult D, Colson P, La Scola B. 2016. Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. Viruses 8:300. https://doi.org/10.3390/v8110300
8. Bajrai LH, Benamar S, Azhar EI, Robert C, Levasseur A, Raoult D, La Scola B. 2016. Kaumoebavirus, a new virus that clusters with Faustoviruses and *Asfarviridae*. Viruses 8:278. https://doi.org/10.3390/v8110278
9. Andreani J, Khalil JYB, Sevvana M, Benamar S, Di Pinto F, Bitam I, Colson P, Klose T, Rossmann MG, Raoult D, La Scola B. 2017. Pacmanvirus, a new giant icosahedral virus at the crossroads between *Asfarviridae* and faustoviruses. J Virol 91:e00212-17. https://doi.org/10.1128/JVI.00212-17
10. Andreani J, Khalil JYB, Baptiste E, Hasni I, Michelle C, Raoult D, Levasseur A, La Scola B. 2017. Orpheovirus IHUMI-LCC2: a new virus among the giant viruses. Front Microbiol 8:2643. https://doi.org/10.3389/fmicb.2017.02643
11. Yoshikawa G, Blanc-Mathieu R, Song C, Kayama Y, Mochizuki T, Murata K, Ogata H, Takemura M. 2019. Medusavirus, a novel large DNA virus discovered from hot spring water. J Virol 93:e02130-18. https://doi.org/10.1128/JVI.02130-18
12. Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat Commun 9:749. https://doi.org/10.1038/s41467-018-03168-1
13. Renesto P, Abergel C, Decloquement P, Moinier D, Azza S, Ogata H, Fourquet P, Gorvel J-P, Claverie J-M. 2006. Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. J Virol 80:11678–11685. https://doi.org/10.1128/JVI.00940-06
14. Fischer MG, Kelly I, Foster LJ, Suttle CA. 2014. The virion of *Cafeteria roenbergensis* virus (CroV) contains a complex suite of proteins for transcription and DNA repair. Virology (Auckl) 466–467:82–94. https://doi.org/10.1016/j.virol.2014.05.029
15. Schrad JR, Abrahão JS, Cortines JR, Parent KN. 2020. Structural and proteomic characterization of the initiation of giant virus infection. Cell 181:1046–1061. https://doi.org/10.1016/j.cell.2020.04.032
16. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M. 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306:1344–1350. https://doi.org/10.1126/science.1101485
17. Machado TB, Picorelli ACR, de Azevedo BL, de Aquino ILM, Queiroz VF, Rodrigues RAL, Araújo JP Jr, Ullmann LS, Dos Santos TM, Marques RE, Guimarães SL, Andrade A, Gularte JS, Demoliner M, Filippi M, Pereira V, Spilki FR, Krupovic M, Aylward FO, Del-Bem L-E, Abrahão JS. 2023. Gene duplication as a major force driving the genome expansion in some giant viruses. J Virol 97:e0130923. https://doi.org/10.1128/jvi.01309-23
18. Bajrai LH, Mougari S, Andreani J, Baptiste E, Delerce J, Raoult D, Azhar EI, La Scola B, Levasseur A. 2019. Isolation of Yasminevirus, the first member of Klosneuvirinae isolated in coculture with *Vermamoeba vermiformis*, demonstrates an extended arsenal of translational apparatus components. J Virol 94:e01534-19. https://doi.org/10.1128/JVI.01534-19
19. Bisio H, Legendre M, Giry C, Philippe N, Alempic JM, Jeudy S, Abergel C. 2023. Evolution of giant pandoravirus revealed by CRISPR/Cas9. Nat Commun 14:428. https://doi.org/10.1038/s41467-023-36145-4

5 – DISCUSSÃO

Em 2003, foi descoberto o primeiro vírus gigante, o APMV (4), e desde então vários outros vírus gigantes vêm sendo descobertos revelando uma grande variedade de partículas, tamanhos de genomas e genes preditos (27,28,25,29,30,31,22,32,26,33). Desde a criação do nosso grupo, em 2011, diferentes técnicas de prospecção foram utilizadas. Ao longo dos anos o grupo foi adaptando e aprimorando cada etapa. A técnica utilizada atualmente tem demonstrado um alto nível de sucesso, sendo possível isolar vírus de várias famílias diferentes (48,49). A prospecção é essencial para a descoberta de novos vírus gigantes pois, assim, mais estudos podem ser realizados, contribuindo para a elucidação de diversas características relacionadas a esses vírus que ainda são um mistério ou pouco compreendidas.

O gigantismo genômico observado nos vírus gigantes é uma característica marcante e intrigante, podendo ser maior até mesmo que de algumas bactérias (4). Embora existam muitos estudos descrevendo o conteúdo funcional dos genomas de vírus gigantes, poucos estudos investigaram os mecanismos envolvidos no aumento do tamanho desses genomas. Estudos apontam que a transferência horizontal de genes, a emergência de genes de novo e a duplicação de genes são mecanismos possivelmente responsáveis por esse gigantismo (27,35,34).

Acredita-se que a duplicação gênica desempenha um papel central nesse processo, sendo desencadeada principalmente por sequências repetitivas presentes no genoma, sendo um mecanismo recorrente em grupos que infectam amebas. Tal duplicação ocorre de diversas formas, como eventos em tandem e rearranjos genômicos distais, frequentemente impulsionados por sequências repetitivas nos genes virais.

A duplicação gênica já era reconhecida como uma importante fonte de diversidade genética em organismos celulares (39,41,42,40). Novas funções podem emergir como resultado de eventos de duplicação, seguidos por divergência (34). Além da divergência funcional potencial, as famílias de genes de múltiplas cópias podem contribuir para o aumento da expressão genética por meio do chamado fenômeno de dosagem genética. Alguns dos exemplos bem conhecidos de famílias de múltiplos genes incluem genes que codificam proteínas formadoras de filamentos citomotores, globinas e unidades ribossômicas (34,50,51). é inegável a importância da duplicação gênica na evolução dos organismos. Entretanto, a compreensão das consequências da duplicação gênica em vírus permanece limitada.

Neste estudo, apresentamos a evidência de que a duplicação gênica é um mecanismo primário para a expansão do genoma entre vários grupos de vírus gigantes. Os genes em cdv.

pambiensis parecem se duplicar por meio de vários mecanismos, envolvendo duplicações de genes em tandem e duplicações de segmentos genômicos distais. Duplicações em tandem em organismos celulares podem ocorrer por meio de deslizamento de replicação, recombinação ectópica ou reparo de quebra de DNA aberrante. Duplicações distais geralmente envolvem rearranjos de crossing-over desigual de grupos de genes com conteúdo genético semelhante. Esse mecanismo pode se tornar progressivamente mais frequente à medida que o número de parálogos aumenta, fornecendo mais regiões com conteúdo semelhante disponíveis para recombinação (38).

Os genes parálogos estão distribuídos por todo o genoma de cedratvirus pambiensis e podem estar em pares, trios ou mesmo compor famílias genicas. Diversas categorias funcionais foram identificadas para os genes duplicados, inclusive algumas potencialmente essenciais, como a principal proteína do capsídeo, fatores de transcrição precoces e enzimas transcricionais.

Todos os bamfordvírus analisados neste estudo apresentaram duplicações gênicas em seus genomas. Além dos cedratvírus, a duplicação de genes também parece ser um fator importante na evolução do genoma dos pandoravírus (média de 47,28%) e alguns mimivírus (para cotonvírus por exemplo, com 49,46%), principalmente. No entanto, não apenas dentre os pitho-like, mas dentre todos os bamfordvírus, o cdv. pambiensis é o vírus com maior percentual de genoma composto por genes duplicados (72,3%). Essas análises indicam que, para alguns vírus, outros mecanismos podem estar atuando simultaneamente, ou talvez sejam os mecanismos principais, em vez da duplicação de genes.

A duplicação gênica é um fenômeno que fornece matéria-prima para a evolução. Manter cópias duplicadas pode ser importante para criar redundância genética, protegendo o vírus de mutações deletérias em genes essenciais, uma vez que cópias adicionais podem manter a funcionalidade e a aptidão do organismo. Pode acontecer também a aquisição de novas funções pelas cópias adicionais, funções não relacionadas às dos genes originais (52,53). Além disso, as regiões genômicas duplicadas fornecem a matéria-prima genética para a emergência de genes de novo, um mecanismo descrito para os pandoravírus (35). Ambos os mecanismos podem levar à inovação genética, aumentando o repertório genético e influenciando no caminho evolutivo desse vírus. Estudos mais aprofundados são necessários para refinar ainda mais nossa compreensão dos mecanismos de expansão do genoma e características evolutivas associadas a este notável grupo de vírus.

Como demonstrado, as análises genômicas são muito importantes para expandir nosso conhecimento sobre os vírus gigantes, contudo as análises proteômicas têm papel fundamental na confirmação dos dados obtidos por bioinformática. Uma característica intrigante dos vírus gigantes é a presença de ORFans em seus genomas, que são genes com funções desconhecidas, mas são frequentemente identificados em análises proteômicas. Estudos proteômicos podem ajudar a elucidar os papéis e a importância dessas proteínas sobre a biologia do vírus, manipulação do ambiente hospedeiro e evolução viral. Apesar do crescente interesse nesses vírus, ainda existem poucos estudos de proteômica sobre eles, o que mantém muitos aspectos de seu ciclo de vida e interação com o hospedeiro ou com o ambiente ainda um mistério.

Neste trabalho foi realizada a proteômica de partículas purificadas de cdv. pambiensis e identificadas 266 proteínas virais. Dentre elas, foram identificadas a presença de proteínas envolvidas em processos como replicação de DNA, transcrição e metabolismo. Essas proteínas já haviam sido descritas para outros vírus gigantes e sugerem que esses vírus possuem certa autonomia em alguns processos metabólicos. A presença de maquinaria completa de transcrição na partícula, por exemplo, indica que o vírus não necessita da maquinaria celular de transcrição, presente no núcleo, conferindo a possibilidade de que o início do ciclo aconteça no citoplasma (27,28,47,45,46,19).

As comparações entre os proteomas de cedratvírus pambiensis e pithovírus sibericum revelaram 89 proteínas compartilhadas. Algumas proteínas revelam uma possível autonomia viral, indicando que esses vírus são capazes de iniciar a replicação e regular sua expressão logo após a penetração no organismo hospedeiro, essas proteínas estão relacionadas, por exemplo, a replicação e reparo do DNA, transcrição e processamento do RNA. Foram identificadas também várias proteínas associadas ao metabolismo de lipídios, nucleotídeos e carboidratos, sugerindo potencial manipulação do ambiente do hospedeiro no início do ciclo.

Foram identificadas também duas proteínas estruturais, a proteína principal do capsídeo (MCP), destacando seus papéis essenciais na montagem do vírion. Além das proteínas virais, 172 proteínas associadas ao hospedeiro *Acanthamoeba castellanii* foram identificadas, indicando a importância dessas proteínas durante a etapa final do ciclo.

Foram identificadas 96 proteínas exclusivas do proteoma do cdv pambiensis, entre elas proteínas envolvidas no metabolismo de nucleotídeos e carboidratos, interações hospedeiro-vírus, estrutura e morfogênese do vírion e tradução.

Esses dados demonstram diferenças significativas quando comparados os proteomas do cedratvirus pambiensis e do pithovirus sibericum. No entanto, algumas proteínas-chave são conservadas, incluindo um conjunto completo de proteínas envolvidas na transcrição, sugerindo que esse vírus pode iniciar e completar seu ciclo de vida sem uma fase nuclear aparente. Mais pesquisas sobre outros pimascovírus são necessárias para obter uma compreensão mais profunda do conjunto essencial de proteínas compartilhadas por esse grupo de vírus.

6 - CONCLUSÕES

- A técnica de prospecção utilizada pelo nosso grupo se mostrou bem sucedida no isolamento de novos vírus gigantes.

- As análises morfológicas do cedratvírus pambiensis, revelaram uma partícula similar à de outros cedratvírus já descritos. Entretanto, foi observado pela primeira vez entre cedratvírus a presença de fibrilas de superfície em seu capsídeo.

- As análises do genoma de cdv. pambiensis, revelaram uma quantidade sem precedentes de genes parálogos em seu genoma (72,3% do genoma composto por genes parálogos).

- Foram identificadas 18 famílias/clusters, além de 17 trios e 126 pares de parálogos no genoma de cdv. pambiensis.

- Os parálogos pertencem a diferentes categorias funcionais, como proteínas contendo domínio de anquirina, proteínas semelhantes à colágeno, proteínas quinases de serina-treonina, proteínas hipotéticas, proteínas contendo domínio F-box, ORFans e outras.

- Para as 6 maiores famílias gênicas foi observada a ocorrência de mais de um tipo de evento de duplicação: duplicações em tandem proximais (principal) e duplicações de segmentos cromossômicos. Já para os pares de parálogos, duplicações em tandem não são frequentes, as duplicações são em sua maioria localizadas distalmente no genoma, provavelmente derivadas de duplicações de segmento cromossômico.

- O fenômeno de pseudogenização foi possivelmente identificado nas 6 maiores famílias gênicas de parálogos no cdv. pambiensis.

- Genes parálogos parecem estar espalhados por todo o genoma de cdv. pambiensis.

- A duplicação gênica e a formação de grandes clusters de parálogos parecem ser universais para todos os membros da família *Pithoviridae*.

- Algumas funções preditas para grandes famílias gênicas de parálogos estão presentes em todos os cedratvírus. Entretanto, para o orpheovírus é observada grande família gênica não observada nos cedratvírus (proteínas com repetição MORN).

- A paralogia está presente em genomas de diferentes representantes de vírus grandes e gigantes do reino *Bamfordvirae*. Entretanto, parece ser mais significativo na expansão do genoma dos pitho-like vírus, alguns *Mimividae* (como o cotonvírus) e pandoravírus.

- O proteoma do cdv. pambiensis possui 266 proteínas virais e 172 proteínas associada ao seu hospedeiro *A. castellanii*.

- Diversas categorias funcionais foram encontradas no proteoma do cdv. pambiensis, inclusive algumas que sugerem um nível importante de autonomia viral, como proteínas relacionadas à replicação e reparo de DNA, transcrição e processamento de RNA.

- Análises comparativas entre os proteomas do cdv. pambiensis e do pithovirus sibericum revelaram que 96 das 266 proteínas identificadas são exclusivas para o cdv. pambiensis, estando ausentes no proteoma de pithovirus sibericum.

- Foram identificadas 89 proteínas em comum nos proteomas de ambos os vírus, incluindo proteínas chaves envolvidas em diversos processos celulares, como replicação, transcrição e metabolismo.

7- REFERÊNCIAS BIBLIOGRÁFICAS

[1] LWOFF, A. The Concept of Virus. **Microbiology**, v. 17, n. 2, p. 239–253, 1957.

[2] KOONIN, E. V.; YUTIN, N. Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. **Intervirology**, v. 53, n. 5, p. 284–292, 15 jun. 2010.

[3] IYER, L. M.; ARAVIND, L.; KOONIN, E. V. Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. **Journal of Virology**, v. 75, n. 23, p. 11720–11734, dez. 2001.

[4] SCOLA, B. L. et al. A Giant Virus in Amoebae. **Science**, v. 299, n. 5615, p. 2033–2033, 28 mar. 2003.

[5] RAOULT, D. et al. The 1.2-Megabase Genome Sequence of Mimivirus. **Science**, v. 306, n. 5700, p. 1344–1350, 19 nov. 2004.

[6] GORBALENYA, A. E. et al. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. **Nature Microbiology**, v. 5, n. 5, p. 668–674, maio 2020.

[7] KOONIN, E. V.; SENKEVICH, T. G.; DOLJA, V. V. The ancient Virus World and evolution of cells. **Biology Direct**, v. 1, n. 1, p. 29, 19 set. 2006.

[8] YUTIN, N. et al. Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. **Virology Journal**, v. 6, n. 1, p. 223, 17 dez. 2009.

[9] COLSON, P. et al. Reclassification of Giant Viruses Composing a Fourth Domain of Life in the New Order Megavirales. **Intervirology**, v. 55, n. 5, p. 321–332, 14 abr. 2012.

[10] COLSON, P. et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. **Archives of Virology**, v. 158, n. 12, p. 2517–2521, 1 dez. 2013.

[11] BOYER, M. et al. Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. **PLOS ONE**, v. 5, n. 12, p. e15530, 2 dez. 2010.

[12] YUTIN, N.; WOLF, Y. I.; KOONIN, E. V. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. **Virology**, Special issue: Giant Viruses. v. 466–467, p. 38–52, 1 out. 2014.

[13] KOONIN, E. V.; YUTIN, N. Multiple evolutionary origins of giant viruses. **F1000Research**, v. 7, p. F1000 Faculty Rev-1840, 22 nov. 2018.

[14] KOONIN, E. V.; YUTIN, N. Chapter Five - Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. Em: KIELIAN, M.; METTENLEITER, T. C.; ROOSSINCK, M. J. (Eds.). **Advances in Virus Research**. [s.l.] Academic Press, 2019. v. 103p. 167–202.

[15] RODRIGUES, R. A. et al. The morphogenesis of different giant viruses as additional evidence for a common origin of *Nucleocytoviricota*. **Current Opinion in Virology**, v. 49, p. 102–110, 1 ago. 2021.

[16] RAOULT, D.; SCOLA, B. L.; BIRTLES, R. The Discovery and Characterization of Mimivirus, the Largest Known Virus and Putative Pneumonia Agent. **Clinical Infectious Diseases**, v. 45, n. 1, p. 95–102, 1 jul. 2007.

[17] ABRAHÃO, J. S. et al. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. **Virology Journal**, v. 11, n. 1, p. 120, 30 jun. 2014.

[18] XIAO, C. et al. Structural Studies of the Giant Mimivirus. **PLOS Biology**, v. 7, n. 4, p. e1000092, 28 abr. 2009.

[19] SCHRAD, J. R. et al. Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. **Cell**, v. 181, n. 5, p. 1046- 1061.e6, 28 maio 2020.
[20] RODRIGUES, R. A. L. et al. Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. **Journal of Virology**, v. 89, n. 23, p. 11812–11819, 5 nov. 2015.
[21] BOYER, M. et al. Mimivirus shows dramatic genome reduction after intraamoebal culture. **Proceedings of the National Academy of Sciences**, v. 108, n. 25, p. 10296–10301, 21 jun. 2011.
[22] ANDREANI, J. et al. Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. **Viruses**, v. 8, n. 11, p. 300, nov. 2016.
[23] SILVA, L. K. DOS S. et al. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. **Scientific Reports**, v. 8, n. 1, p. 4000, 5 mar. 2018.
[24] RODRIGUES, R. A. L. et al. Morphologic and Genomic Analyses of New Isolates Reveal a Second Lineage of Cedratviruses. **Journal of Virology**, v. 92, n. 13, p. 10.1128/jvi.00372-18, 13 jun. 2018.
[25] LEGENDRE, M. et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. **Proceedings of the National Academy of Sciences**, v. 111, n. 11, p. 4274–4279, 18 mar. 2014.
[26] ANDREANI, J. et al. Orpheovirus IHUMI-LCC2: A New Virus among the Giant Viruses. **Frontiers in Microbiology**, v. 8, 22 jan. 2018.
[27] BOYER, M. et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. **Proceedings of the National Academy of Sciences**, v. 106, n. 51, p. 21848–21853, 22 dez. 2009.
[28] PHILIPPE, N. et al. Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. **Science**, v. 341, n. 6143, p. 281–286, 19 jul. 2013.
[29] RETENO, D. G. et al. Faustovirus, an Asfarvirus-Related New Lineage of Giant Viruses Infecting Amoebae. **Journal of Virology**, v. 89, n. 13, p. 6585–6594, 3 jun. 2015.
[30] LEGENDRE, M. et al. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. **Proceedings of the National Academy of Sciences**, v. 112, n. 38, p. E5327–E5335, 22 set. 2015.
[31] BAJRAI, L. H. et al. Kaumoebavirus, a New Virus That Clusters with Faustoviruses and Asfarviridae. **Viruses**, v. 8, n. 11, p. 278, nov. 2016.
[32] ANDREANI, J. et al. Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. **Journal of Virology**, v. 91, n. 14, p. 10.1128/jvi.00212-17, 26 jun. 2017.
[33] YOSHIKAWA, G. et al. Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. **Journal of Virology**, v. 93, n. 8, p. 10.1128/jvi.02130-18, 3 abr. 2019.
[34] SUHRE, K. Gene and Genome Duplication in Acanthamoeba polyphaga Mimivirus. **Journal of Virology**, v. 79, n. 22, p. 14095–14101, 15 nov. 2005.
[35] LEGENDRE, M. et al. Diversity and evolution of the emerging Pandoraviridae family. **Nature Communications**, v. 9, n. 1, p. 2285, 11 jun. 2018.
[36] OCHMAN, H.; LAWRENCE, J. G.; GROISMAN, E. A. Lateral gene transfer and the nature of bacterial innovation. **Nature**, v. 405, n. 6784, p. 299–304, maio 2000.
[37] OLIVEIRA, G. P. et al. Promoter Motifs in NCLDVs: An Evolutionary Perspective. **Viruses**, v. 9, n. 1, p. 16, jan. 2017.
[38] KREBS, J. E. **Lewin's genes XII**. Burlington, MA: Jones & Bartlett Learning, 2018.
[39] CANNON, S. B. et al. The roles of segmental and tandem gene duplication in the

evolution of large gene families in Arabidopsis thaliana. **BMC Plant Biology**, v. 4, n. 1, p. 10, 1 jun. 2004.

[40] PERSI, E. et al. Compensatory relationship between low-complexity regions and gene paralogy in the evolution of prokaryotes. **Proceedings of the National Academy of Sciences**, v. 120, n. 16, p. e2300154120, 18 abr. 2023.

[41] REAMS, A. B.; NEIDLE, E. L. Selection for Gene Clustering by Tandem Duplication. **Annual Review of Microbiology**, v. 58, n. Volume 58, 2004, p. 119–142, 13 out. 2004.

[42] WAPINSKI, I. et al. Natural history and evolutionary principles of gene duplication in fungi. **Nature**, v. 449, n. 7158, p. 54–61, set. 2007.

[43] LI, W.-H. et al. Evolutionary analyses of the human genome. **Nature**, v. 409, n. 6822, p. 847–849, fev. 2001.

[44] ZHANG, J. Evolution by gene duplication: an update. **Trends in Ecology & Evolution**, v. 18, n. 6, p. 292–298, 1 jun. 2003.

[45] RENESTO, P. et al. Mimivirus Giant Particles Incorporate a Large Fraction of Anonymous and Unique Gene Products. **Journal of Virology**, v. 80, n. 23, p. 11678–11685, dez. 2006.

[46] FISCHER, M. G. et al. The virion of *Cafeteria roenbergensis* virus (CroV) contains a complex suite of proteins for transcription and DNA repair. **Virology**, Special issue: Giant Viruses. v. 466–467, p. 82–94, 1 out. 2014.

[47] ABRAHÃO, J. et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. **Nature Communications**, v. 9, n. 1, p. 749, 27 fev. 2018.

[48] ANDRADE, A. C. DOS S. P. et al. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. **Virology Journal**, v. 15, n. 1, p. 22, 24 jan. 2018.

[49] MACHADO, T. B. et al. A long-term prospecting study on giant viruses in terrestrial and marine Brazilian biomes. **Virology Journal**, v. 21, n. 1, p. 135, 10 jun. 2024.

[50] WICKSTEAD, B.; GULL, K. The evolution of the cytoskeleton. **Journal of Cell Biology**, v. 194, n. 4, p. 513–525, 22 ago. 2011.

[51] MALIK GHULAM, M. et al. Duplicated ribosomal protein paralogs promote alternative translation and drug resistance. **Nature Communications**, v. 13, n. 1, p. 4938, 23 ago. 2022.

[52] KOONIN, E. V.; DOLJA, V. V.; KRUPOVIC, M. The logic of virus evolution. **Cell Host & Microbe**, v. 30, n. 7, p. 917–929, 13 jul. 2022.

[53] KOONIN, E. V.; KRUPOVIC, M. The depths of virus exaptation. **Current Opinion in Virology**, Virus structure and expression • Viral evolution. v. 31, p. 1–8, 1 ago. 2018.

8 - EVENTOS CIENTÍFICOS E PRODUÇÕES

8.1 - Participações em eventos científicos

- Participação no XXXII Congresso Brasileiro de Virologia e XVI Encontro de Virologia do Mercosul, 2021 - Virologia em Casa, Online
- Participação e apresentação de pôster no VIII Simpósio de Microbiologia da UFMG: CONECTA SIM 2021, Online.
- Participação e apresentação oral no XXXIII Congresso Brasileiro de Virologia 2022 e XVII Encontro de Virologia do Mercosul, 2022, Bahia.
- Participação a apresentação de pôster no IX Simpósio de Microbiologia da UFMG, 2022
  - Recebi prêmio de pôster com relevância acadêmica
- Participação no Curso de Inverno em Bioinformática 2023 da UFMG, 2023, Belo Horizonte.
- Participação e apresentação oral no V FAMERP- UTMB: Emerging infections in the Americas- Common interests and collaboration between North-South, 2023, São José do Rio Preto/SP.
- Participação e pôster no XXXIV Congresso Brasileiro de Virologia e XVIII Encontro de Virologia do Mercosul, 2023, Ouro Preto.
- Participação e apresentação de pôster no CeZAP 2024 Infectious Disease Symposium, Virginia Tech, EUA.

8.2 - Formação complementar

- Participação no Minicurso *Filogenia Viral* (4h) no Curso de Inverno em Bioinformática 2023 da UFMG, 2023, Belo Horizonte
- Participação no Minicurso *Tá tudo sobre controle: Versionamento de código com Git e GitHub* (4h) no Curso de Inverno em Bioinformática 2023 da UFMG, 2023, Belo Horizonte
- Realização do Doutorado Sanduíche na Virginia Tech nos Estados Unidos. Duração de 9 meses, sob orientação do Professor Frank Aylward.

8.3 - Atividades de extensão

- Participação como analista de laboratório voluntária no enfrentamento a pandemia de coronavírus pela iniciativa Rede de Laboratórios – UFMG, 2021, Laboratório de Vírus, Belo Horizonte.
- Participação em uma expedição no Pantanal realizada pela Fiocruz em parceria com a Marinha do Brasil para coleta de amostras e auxiliar nas análises realizadas pelo Projeto NAVIO (Navegação Ampliada para Vigilância Intensiva e Otimizada), 2023.

8.4 – Coorientação

- Coorientação do aluno de Iniciação Científica e graduando em Ciências Biológicas pela UFMG, Matheus Gomes Barcelos. Projeto intitulado: Prospecção de Vírus Gigantes em Biomas Brasileiros. Período: 2022.
- Coorientação do aluno de Mestrado Matheus Gomes Barcelos. Projeto intitulado: Isolamento e Caracterização Biológica de Microrganismos Associados a Amebas em Amostras Ambientais. Período: 2023-2024

8.5 - Participação em bancas

- Participação em banca de defesa de Trabalho de Conclusão de Curso (Graduação em Ciências Biológicas) de Clécio Alonso da Costa Filho. Trabalho intitulado: Explorando o genoma de vírus grandes de algas para fins biotecnológicos: caracterização genômica funcional e filogenética de chlorovírus. Em 2023 na Universidade Federal de Minas Gerais, Belo Horizonte.

8.6 - Artigos científicos publicados durante o período da Tese

1. **MACHADO, TALITA BASTOS**; AQUINO, ISABELLA LUIZA MARTINS; ABRAHÃO, JÔNATAS SANTOS. Isolation of Giant Viruses of *Acanthamoeba castellanii*. Current Protocols, v. 2, p. 1, 2022.

2. BORATTO, PAULO VICTOR M. ; SERAFIM, MATEUS SÁ M. ; WITT, AMANDA STÉPHANIE A. ; CRISPIM, ANA PAULA C. ; AZEVEDO, BRUNA LUIZA DE ; SOUZA, GABRIEL AUGUSTO P. DE ; AQUINO, ISABELLA LUIZA M. DE ; **MACHADO, TALITA B**. ; QUEIROZ, VICTÓRIA F. ; RODRIGUES, RODRIGO A. L. ; BERGIER, IVAN ; CORTINES, JULIANA REIS ; FARIAS, SAVIO TORRES DE ; SANTOS, RAÍSSA NUNES DOS ; CAMPOS, FABRÍCIO SOUZA ; FRANCO, ANA CLÁUDIA ; ABRAHÃO, JÔNATAS S. A Brief History of Giant Viruses? Studies in Brazilian Biomes. Viruses-Basel, v. 14, p. 191, 2022

3. DE AQUINO, ISABELLA LUIZA MARTINS; SERAFIM, MATEUS SÁ MAGALHÃES; **MACHADO, TALITA BASTOS**; AZEVEDO, BRUNA LUIZA; CUNHA, DENILSON EDUARDO SILVA; ULLMANN, LEILA SABRINA; ARAÚJO, JOÃO PESSOA; ABRAHÃO, JÔNATAS SANTOS. Diversity of Surface Fibril Patterns in Mimivirus Isolates. JOURNAL OF VIROLOGY, v. 97, p. e0182422, 2023.

4. AQUINO, ISABELLA LUIZA MARTINS DE; BARCELOS, MATHEUS GOMES; **MACHADO, TALITA BASTOS**; SERAFIM, MATEUS SÁ MAGALHÃES; ABRAHÃO, JÔNATAS SANTOS. Surface fibrils on the particles of nucleocytoviruses: A review. Experimental Biology and Medicine, p. 2045-2052, 2023.

5. **MACHADO, TALITA B.**; PICORELLI, AGNELLO C. R. ; DE AZEVEDO, BRUNA L. ; DE AQUINO, ISABELLA L. M. ; QUEIROZ, VICTÓRIA F. ; RODRIGUES, RODRIGO A. L. ; ARAÚJO, JOÃO PESSOA ; ULLMANN, LEILA S. ; DOS SANTOS, THIAGO M. ; MARQUES, RAFAEL E. ; GUIMARÃES, SAMUEL L. ; ANDRADE, ANA CLÁUDIA S. P. ; GULARTE, JULIANA S. ; DEMOLINER, MERIANE ; FILIPPI, MICHELI ; PEREIRA, VYCTORIA M. A. G. ; SPILKI, FERNANDO R. ; KRUPOVIC, MART ; AYLWARD, FRANK O. ; DEL-BEM, LUIZ-EDUARDO. Gene duplication as a major force driving the genome expansion in some giant viruses. JOURNAL OF VIROLOGY, v. 97, p. 1, 2023.

6. **MACHADO, TALITA B**.; DE AQUINO, ISABELLA L. M.; AZEVEDO, BRUNA L.; SERAFIM, MATEUS S.; BARCELOS, MATHEUS G.; ANDRADE, ANA CLÁUDIA S. P.; REIS, ERIK; ULLMANN, LEILA SABRINA; PESSOA, JOÃO; COSTA, ADRIANA O.; ROSA, LUIZ H.; ABRAHÃO, JÔNATAS S. A long-term prospecting study on giant viruses in terrestrial and marine Brazilian biomes. Virology Journal, v. 21, p. 135, 2024.

7. DE AZEVEDO, BRUNA LUIZA; QUEIROZ, VICTÓRIA FULGÊNCIO; DE AQUINO, ISABELLA LUIZA MARTINS; **MACHADO, TALITA BASTOS**; DE ASSIS, FELIPE LOPES; REIS, ERIK; ARAÚJO JÚNIOR, JOÃO PESSOA; ULLMANN, LEILA SABRINA; COLSON, PHILIPPE; GREUB, GILBERT; AYLWARD, FRANK; RODRIGUES, RODRIGO ARAÚJO LIMA; ABRAHÃO, JÔNATAS SANTOS. The genomic and phylogenetic analysis of Marseillevirus cajuinensis raises questions about the evolution of *Marseilleviridae* lineages and their taxonomical organization. JOURNAL OF VIROLOGY, v. 98, p. e0051324, 2024.

8.7 – Artigo científico em processo final de escrita

1. **MACHADO, TALITA B**.; HANCHUK, TALITA D. M.; DE AQUINO, ISABELLA L. M.; AZEVEDO, BRUNA L.; SILVA, RAFAEL E. M. P.; ABRAHÃO, JÔNATAS SANTOS. The proteomics of the giant cedratvirus particles reveals unique and shared features with pitho-like viruses. Será submetido no periódico Journal of Virology.

8.8 – Artigos científicos publicados anteriormente

1. RODRIGUES, RODRIGO ARAÚJO LIMA ; ANDREANI, JULIEN ; ANDRADE, ANA CLÁUDIA DOS SANTOS PEREIRA ; **MACHADO, TALITA BASTOS** ; ABDI, SOUHILA ; LEVASSEUR, ANTHONY ; ABRAHÃO, JÔNATAS SANTOS ; LA SCOLA, BERNARD . Morphologic and Genomic Analyses of New Isolates Reveal a Second Lineage of Cedratviruses. JOURNAL OF VIROLOGY, v. 92, p. 29695424, 2018.

2. ANDRADE, ANA CLÁUDIA DOS SANTOS PEREIRA ; BORATTO, PAULO VICTOR DE MIRANDA ; RODRIGUES, RODRIGO ARAÚJO LIMA ; **MACHADO, TALITA BASTOS**; AZEVEDO, BRUNA LUIZA ; DORNAS, FÁBIO PIO ; OLIVEIRA, DANILO BRETAS ; DRUMOND, BETÂNIA PAIVA ; KROON, ERNA GEESSIEN ; ABRAHÃO, JÔNATAS SANTOS . New isolates of pandoraviruses: contribution to the study of replication cycle steps. JOURNAL OF VIROLOGY, v. 1, p. 01942-18, 2018.

3. ANDRADE, ANA CLÁUDIA DOS SANTOS PEREIRA. ; ARANTES, THALITA S. ; RODRIGUES, RODRIGO A. L. ; **MACHADO, TALITA B**. ; DORNAS, FÁBIO P. ; LANDELL, MELISSA F. ; FURST, CINTHIA ; BORGES, LUIZ G. A. ; DUTRA, LARA A. L. ; ALMEIDA, GABRIEL ; TRINDADE, GILIANE DE S. ; BERGIER, IVAN ; ABRAHÃO, WALTER ; BORGES, IARA A. ; CORTINES, JULIANA R. ; DE OLIVEIRA, DANILO B. ; KROON, ERNA G. ; ABRAHÃO, JÔNATAS S. . Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. Virology Journal, v. 15, p. 1-10, 2018.

4. ABRAHÃO, JÔNATAS SANTOS ; SACCHETTO, LÍVIA ; REZENDE, IZABELA MAURÍCIO ; RODRIGUES, RODRIGO ARAÚJO LIMA ; CRISPIM, ANA PAULA CORREIA ; MOURA, CÉSAR ; MENDONÇA, DIOGO CORREA ; REIS, ERIK ; SOUZA, FERNANDA ; OLIVEIRA, GABRIELA FERNANDA GARCIA ; DOMINGOS, IAGO ; DE MIRANDA BORATTO, PAULO VICTOR ; SILVA, PEDRO HENRIQUE BASTOS ; QUEIROZ, VICTORIA FULGÊNCIO ; **MACHADO, TALITA BASTOS** ; ANDRADE, LUIS ADAN FLORES ; LOURENÇO, KARINE LIMA ; SILVA, THAÍS ; OLIVEIRA, GRAZIELE PEREIRA ; DE SOUZA ALVES, VIVIANE . Detection of SARS-CoV-2 RNA on public surfaces in a densely populated urban area of Brazil: A potential tool for monitoring the circulation of infected patients. SCIENCE OF THE TOTAL ENVIRONMENT, v. 766, p. 142645, 2020.

5. BORATTO, PAULO V. M. ; OLIVEIRA, GRAZIELE P. ; **MACHADO, TALITA B.** ; ANDRADE, ANA CLÁUDIA S. P. ; BAUDOIN, JEAN-PIERRE ; KLOSE, THOMAS ; SCHULZ, FREDERIK ; AZZA, SAÏD ; DECLOQUEMENT, PHILIPPE ; CHABRIÈRE, ERIC ; COLSON, PHILIPPE ; LEVASSEUR, ANTHONY ; LA SCOLA, BERNARD ; ABRAHÃO, JÔNATAS S. . Yaravirus: A novel 80-nm virus infecting *Acanthamoeba castellanii*. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA (ONLINE), v. 117, p. 202001637-16586, 2020.

ANEXO – OUTROS ARTIGOS PUBLICADOS DURANTE O DOUTORADO

Machado *et al. Virology Journal* (2024) 21:135
https://doi.org/10.1186/s12985-024-02404-z

Virology Journal

RESEARCH Open Access

A long-term prospecting study on giant viruses in terrestrial and marine Brazilian biomes

Talita B. Machado[1], Isabella L. M. de Aquino[1], Bruna L. Azevedo[1], Mateus S. Serafim[1], Matheus G. Barcelos[1], Ana Cláudia S. P. Andrade[2], Erik Reis[3], Leila Sabrina Ullmann[4], João Pessoa Jr.[5], Adriana O. Costa[6], Luiz H. Rosa[7] and Jônatas S. Abrahão[1*]

Abstract

The discovery of mimivirus in 2003 prompted the search for novel giant viruses worldwide. Despite increasing interest, the diversity and distribution of giant viruses is barely known. Here, we present data from a 2012–2022 study aimed at prospecting for amoebal viruses in water, soil, mud, and sewage samples across Brazilian biomes, using *Acanthamoeba castellanii* for isolation. A total of 881 aliquots from 187 samples covering terrestrial and marine Brazilian biomes were processed. Electron microscopy and PCR were used to identify the obtained isolates. Sixty-seven amoebal viruses were isolated, including mimiviruses, marseilleviruses, pandoraviruses, cedratviruses, and yaraviruses. Viruses were isolated from all tested sample types and almost all biomes. In comparison to other similar studies, our work isolated a substantial number of Marseillevirus and cedratvirus representatives. Taken together, our results used a combination of isolation techniques with microscopy, PCR, and sequencing and put highlight on richness of giant virus present in different terrestrial and marine Brazilian biomes.

Keywords Giant virus, Amoebas, Biomes, Diversity, Richness

*Correspondence:
Jônatas S. Abrahão
jonatas.abrahao@gmail.com; jsa@icb.ufmg.br
[1]Laboratório de vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte city, Minas Gerais, Brazil
[2]Centre de Recherche du Centre Hospitalier Universitaire de Québec, Université Laval, Laval city, Québec, Canada
[3]Laboratório de virologia básica e aplicada, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte city, Minas Gerais, Brazil
[4]Laboratório de Virologia Veterinária, Faculdade de Medicina Veterinária e Zootecnia (FAMEZ), Universidade Federal de Mato Grosso do Sul (UFMS), Campo Grande city, Mato Grosso do Sul, Brazil
[5]Laboratório de Virologia, Departamento de Microbiologia e Imunologia, Instituto de Biotecnologia, Universidade Estadual Paulista (UNESP), Botucatu city, São Paulo, Brazil
[6]Departamento de Análises Clínicas e Toxicológicas, Faculdade de Farmácia, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte city, Minas Gerais, Brazil
[7]Laboratório de Microbiologia Polar e Conexões Tropicais, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte city, Minas Gerais, Brazil

Introduction

Amoeba giant viruses are well-known for their structural and genomic complexity. Since the mimivirus discovery in 2003 [1], several groups of giant viruses have been described. Metagenomics and prospective studies involving virus isolation revealed that giant amoeba viruses are distributed in a diverse range of environments and substrates [2–8]. These entities have been already isolated from water samples, animals' bodies, permafrost, ocean depths, thermal springs, and soda lakes [9–12].

Brazil is one of the countries with the highest biodiversity and species richness in the world [13]. Its tropical forests, such as the Amazon, are globally renowned for their complexity and their role in regulating the world's climate. However, tropical forests represent just a part of the vast Brazilian territory. Brazil is described as having six major biomes: Amazon (i) and Atlantic Forest (ii), both typical of humid tropical forests and biodiversity hotspots; Cerrado (iii), a savanna biome with a rich network of rivers and a high number of endemic species; Caatinga (iv), an arid biome with species well-adapted to water scarcity; Pantanal (v), one of the world's largest wetland plains; and Pampas (vi), a biome typical of the cooler regions of Brazil, composed of grasslands [14]. In this context, our group has been prospecting viruses of amoebas in Brazilian territory, and we have described completely new species, including Tupanvirus and Yaravirus [11, 15].

In this present work, we describe our efforts in the prospective study of giant viruses from 2012 to 2022. A total of 67 viruses were isolated and identified, from almost all Brazilian biomes. Our study provides information into the isolation and richness of giant amoeba viruses across Brazilian territory, shedding light on their presence in both natural and urban environments.

Materials and methods

Cells and medium

The free-living amoeba of the species *Acanthamoeba castellanii* from the American Type Culture Collection (ATCC 30,234; Maryland, USA) was used in all experiments. The amoebas were propagated in cell culture flasks using peptone-yeast extract-glucose medium (PYG) supplemented with 100 IU/mL of penicillin (Cellofarm, Brazil), 0.25 µg/mL of amphotericin B (Cultilab, Brazil), and 0.1 mg/mL of streptomycin (Sigma-Aldrich, USA). To perform the subcultures, the amoebas were mechanically detached from the monolayer by tapping the bottle, quantified in a Neubauer chamber, and the necessary amount was inoculated into a new flask with fresh medium.

Samples collection

The samples were collected from 2012 to 2022. It is important to note that this study does not intend to test a similar number of samples per biome, state, or substrate. This is a descriptive work on the richness of giant viruses found in Brazil in a 11-years study. Brazil has a vast territory, and reaching some regions can be challenging due to logistical or funding issues. Therefore, several collections prospected here were donated to our group by local researchers. The collections were carried out with authorization from the Biodiversity Authorization and Information System (SISBIO) under the identification numbers: 33326-2 and 34293-2; and SISGEN: A2291C9. Most of the collections were made using mainly sterile 1.5 mL microtubes or 50 mL conical tubes. After collection, the samples underwent three cycles of freezing and thawing (a process to reduce contamination) and were organized into collections. For this purpose, each sample was aliquoted into 1.5 mL microtubes, with 1 to 10 aliquots of 1 mL, depending on the initial quantity of the sample, and stored at -20 °C. At the end of the organization, 24 collections were obtained, collected from different Brazilian states and biomes (Fig. 1A). Samples were obtained from different Brazilian states: Amazonas, Bahia, Goiás, Maranhão, Mato Grosso do Sul, Minas Gerais, Piauí, Rio Grande do Sul, Santa Catarina, and São Paulo. These samples resulted in 1 collection from the Pantanal biome, 1 collection from the Caatinga biome, 1 collection from the Amazon biome, 1 collection from the Pampa biome, 6 collections from the Atlantic Forest biome, 13 collections from the Cerrado biome (Sup. Table 1). At last, 61 samples were also collected from the Atlantic Ocean inside the Brazilian Exclusive Economic Zone (EEZ) from Rio de Janeiro to Rio Grande do Norte, corresponding to 1 collection. From those 24 collections, a total of 881 aliquots of 187 samples were tested for giant viruses. The majority of the samples originated from freshwater (459 aliquots), followed by sewage (110 aliquots), saltwater (300 aliquots), mud (10 aliquots), and soil (2 aliquots) (Fig. 1B).

Prospection of giant viruses in *Acanthamoeba castellanii*

For prospection of giant viruses, preliminary processing of the samples is carried out. For water and sewage samples, a dilution was performed at a ratio of 1:10 for each aliquot. For mud samples, it was necessary to wait 24 h at 10 °C for the sediment to settle first and then perform a dilution at a ratio of 1:10 for each aliquot with the liquid content of the supernatant. For soil samples, 100 g of sediment was collected and 300 µL of sterile distilled water was added, resulting in an initial dilution of 1:4; this content was then vortexed, and the sample was left in the refrigerator overnight for sediment settling. After this period, a dilution at a ratio of 1:10 for each aliquot was

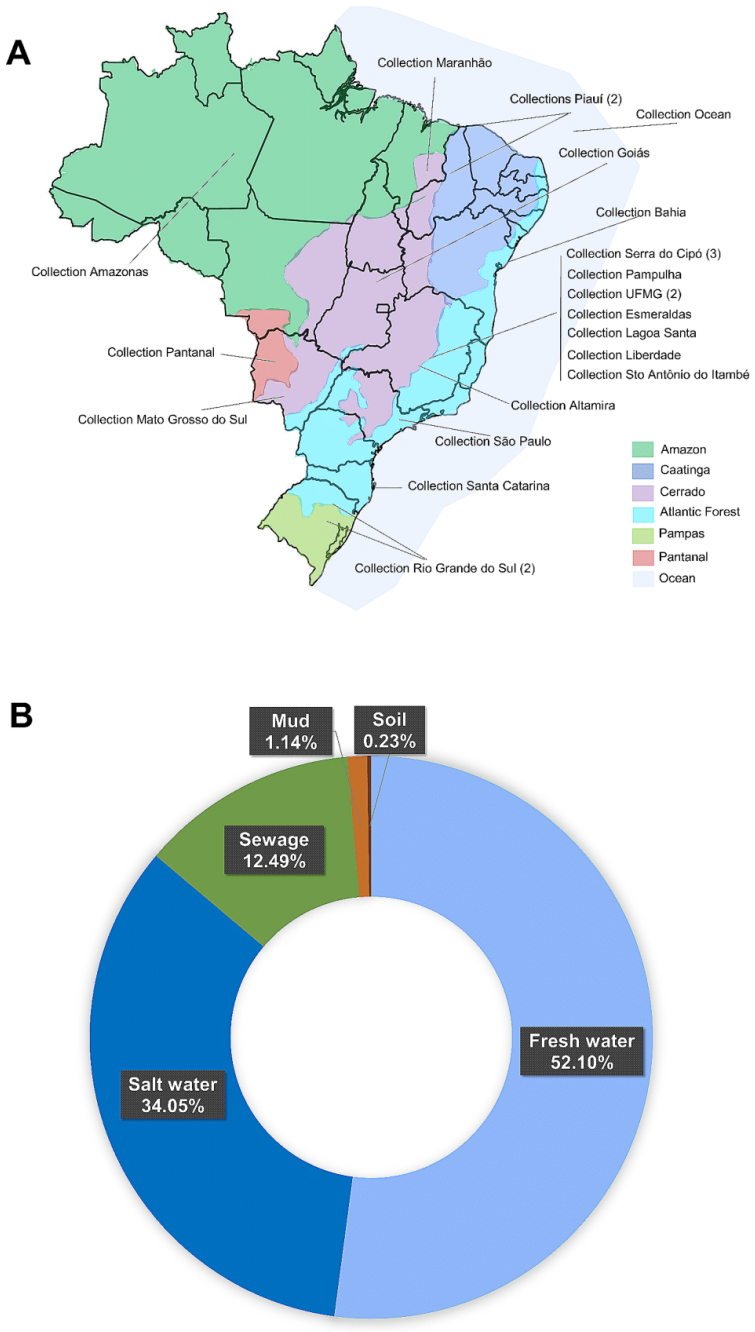

Fig. 1 (**A**) Map with the distribution of collections. Location of each collection based on its geographical coordinates. Brazilian biomes are represented by different colors, as per the figure legend. (**B**) Types of samples considering all the collections

performed with the liquid content of the supernatant. The 1:10 dilutions were made using 1x phosphate-buffered saline (PBS). Virus isolation was performed in *Acanthamoeba castellanii* seeded in 96-well plates (Kasvi, Brazil) is PYG medium supplemented with three additional antibiotics: 0.004 mg/mL vancomycin (Inlab, Brazil), 0.004 mg/mL ciprofloxacin (Sigma-Aldrich, USA), and 0.020 mg/mL doxycycline (Sigma-Aldrich, USA). Inoculated amoebas were monitored for seven days, during three rounds of blind passages. In case of cytopathic effects, the extract of amoebas was collected and sent to virus identification [16].

Transmission electron microscopy

For transmission electron microscopy (TEM), 1.5×10^7 cells of *A. castellanii* were added to T-75 cell culture flasks (Kasvi, Brazil). The cells were infected with the newly isolated viruses during this work at a multiplicity of infection (MOI) of 0.01 and maintained in incubators at a temperature of approximately 30 °C until the cytopathic effect appearance. After observing the cytopathic effect, the contents of the bottle were collected into a 50 mL conical tube and centrifuged at 1308 x g (Sorvall RT6000B) for 10 min. Following centrifugation, the supernatant was discarded, and the pellet was washed with 5 mL of 0.1 M monosodium phosphate buffer, and the contents were transferred to a 15 mL conical tube and centrifuged again at 1308 x g (Sorvall RT6000B) for 10 min, washing and centrifugation were performed twice. After washing, the supernatant was discarded, and the pellet was resuspended in 1.5 mL of fixative composed of 2.5% glutaraldehyde and 0.1 M monosodium phosphate buffer and kept in a homogenizer for 2 h at room temperature. After the incubation period with the fixative, the contents of the conical tube were centrifuged at 145 x g (Sorvall RT6000B) for 10 min, the supernatant was discarded, and the pellet was resuspended with 1 mL of 0.1 M monosodium phosphate buffer, the contents were transferred to a 1.5 mL microtube and centrifuged again at 0.8 x g (Eppendorf 5415R) for 10 min. After this final centrifugation, the material was correctly identified and stored in the refrigerator until it was sent to the Microscopy Center of UFMG. There, it underwent subsequent fixation with 2% osmium tetroxide, embedding in EPON resin, and preparation of ultra-thin sections. Image analysis was performed using a transmission electron microscope (FEI SpiritBiotwin 120 kV). The identification of the viruses was performed based on multiple features from the particles (morphology, size, unique structures and capsid characteristics) combined with impact of infection on the host (viral factory appearance and reorganization of host cytoplasm).

PCR identification and sequencing

Following the identification of cultures exhibiting a cytopathic effect, screening was conducted via PCR targeting specific giant virus groups (Sup. Table 2). DNA extraction was performed using the phenol-chloroform method from 200 µL of the collected content from positive aliquots, yielding DNA at a concentration of approximately 50 µg/µl, utilized as a template for PCR assays. PCR assays targeted various genes, including the major capsid protein gene of mimivirus, Marseillevirus and yaravirus; and the DNA polymerase gene of Pandoravirus, pithovirus and cedratvirus. Design and standardization of primers and reactions were ensured to prevent cross-amplification among analyzed viruses available on GenBank.

PCR assays utilized 1 µL of extracted DNA (~50 nanograms) in an amplification reaction mix containing 5 µL of SYBR Green Master Mix (Thermo Fisher Scientific, USA) and 0.4 µL (10 µM) of forward and reverse primers, adjusted with ultrapure water to a final volume of 10 µL. Thermal cycling conditions on the StepOne thermal cycler (Applied Biosystem, USA) comprised initial denaturation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min, with a final step of 95 °C for 15 s, 60 °C for 1 min, and 95 °C for 15 s. Positive samples exhibited amplification with specific melting temperatures, while negative samples showed no specific amplification. As negative controls, DNA extracted from non-inoculated amoebas with purified viruses or samples was used, while DNA from amoebae infected with purified virus served as a positive control.

Phylogenetic analyses

Some representative viruses of each group were sequenced. The samples containing purified virus underwent sequencing using the Illumina MiSeq system, employing a paired-end library and an Illumina DNA Prep kit (Illumina Inc., San Diego, CA, USA). Quality control of the obtained reads was conducted using the FastQC program, followed by read trimming with the Trimmomatic tool. Genome de novo assembly was performed using Spades 3.12 with default parameters [17].

Phylogenetic trees were constructed using IQtree software (version 1.6.12) employing the maximum-likelihood method, with 1,000 bootstrap replicates for branch support [18]. For tree construction, sequences of different genes were utilized. Data sets containing sequences for alignments were prepared using BLASTp against the NCBI non-redundant protein sequences (nr) database with an expected threshold of 10^{-3} [19]. Alignments were performed using the Muscle software 3.8.1551 [20]. The best-fit substitution models were selected using the ModelFinder algorithm implemented in IQtree. Visualization

and editing of phylogenetic trees were carried out using iTOL.

Results

In this prospective study, covering all Brazilian biomes, a total of 67 amoeba-associated viruses were isolated (Fig. 2). These viruses induced rounding and lysis as cytopathic effects in *Acanthamoeba*. Additionally, one of the isolates caused cell aggregation (Sup. Table 3). Only four isolates were obtained during the 1st passage, 11 isolates during the 2nd passage and the others (52) were detected during the 3rd passage of the prospecting procedures (Sup. Table 3). In addition, no isolates were obtained from the Amazon and Atlantic Ocean samples.

All isolates were first submitted to transmission electron microscopy (TEM) for identification (Figs. 2, 3 and 4). Here we will describe the types of isolated viruses and the criteria we used to identify them by TEM:

−1 Pandoravirus isolate, with its typical ovoid shape particle and approximately 1 µm in length. It presents an apical ostiole, serving as an entry point for interactions with host cells. Within the host cell's cytoplasm, an electron-lucent viral factory is discernible, indicative of active viral replication and assembly processes taking place (Figs. 2 and 3F).

−13 mimiviruses, in which the viral structure exhibits pseudo-icosahedral symmetry, with a diameter measuring approximately 450 nm. Surrounding the capsid is a layer of fibrils, measuring approximately 125 nm in length. The capsid consists of multiple layers of protein. Enclosed within is an internal lipid membrane. Notably, the presence of the "stargate" feature. Within the host cell's cytoplasm, an electron-dense viral factory is discernible, indicative of active viral replication and assembly processes (Figs. 2, 3A and 4A).

−26 marseilleviruses, with icosahedral symmetry, these viral particles have a diameter of 180–250 nm. They may be observed individually or clustered together, inside vesicles. Within the host cell's cytoplasm, an electron-luminous viral factory is evident, indicating active viral replication and assembly processes (Figs. 2, 3B and 4C and D).

−24 pitho/cedrat-like viruses: The viral structure presents an ovoid shape, ranging from 600 nm to 1.5 µm in length. Its capsid is characterized by parallel vertical striations. Notably, one or two apical corks, also striated, can be observed. By electron microscopy, it is not possible to be sure if the isolate is a cedrat- or pithovirus, because particles with one or two corks have already been described for both groups of viruses. Within the

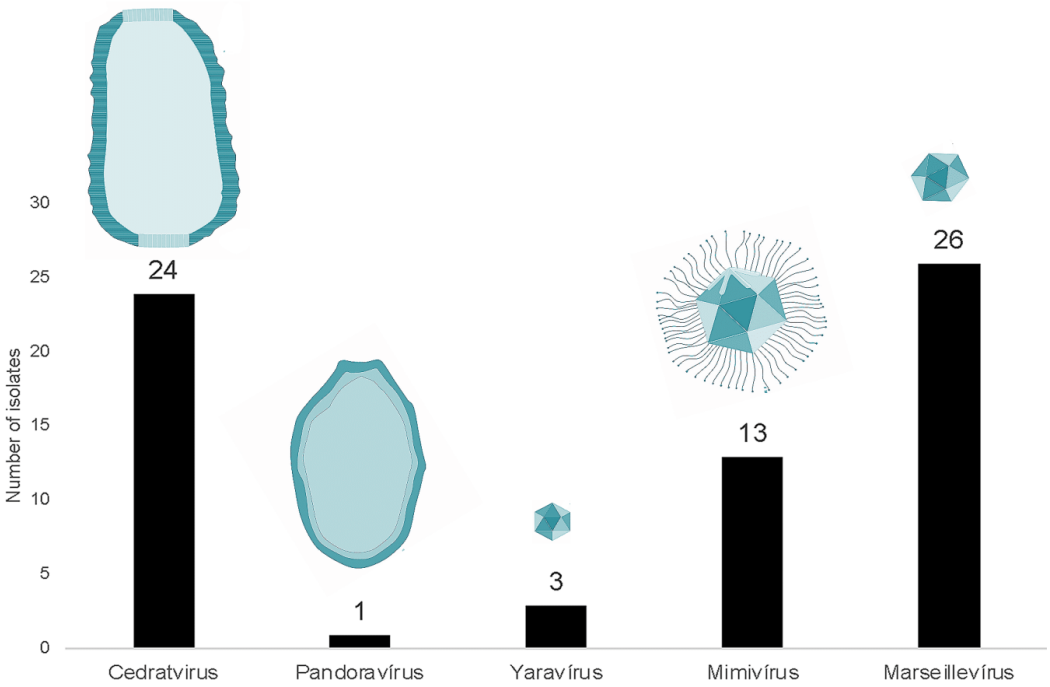

Fig. 2 Number and variety of viruses isolated during this study. A total of 67 isolates were obtained from Brazilian biomes

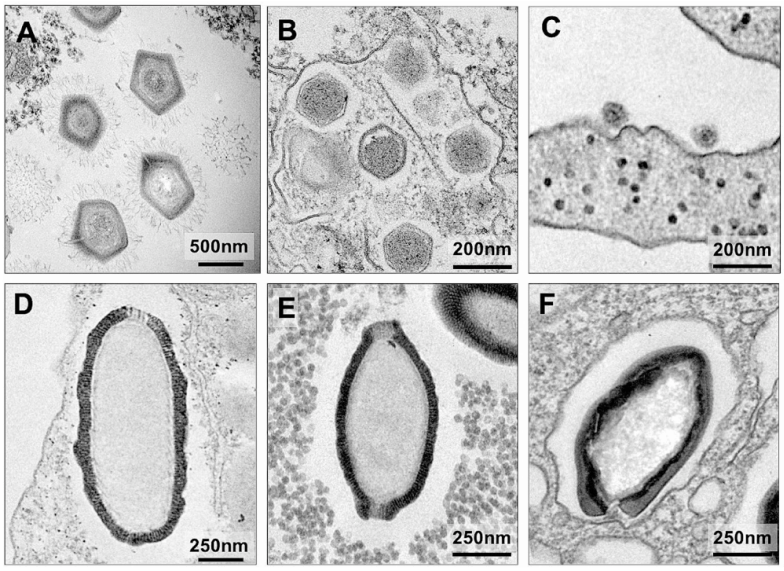

Fig. 3 Transmission electron microscopy of isolated viral particles. (**A**) Mimivirus particles exhibiting its surface fibrils. (**B**) Marseillevirus particles with their typical icosahedral symmetry. (**C**) Yaravirus particles attached to the *Acanthamoeba castellanii* plasma membrane. (**D**) Cedratvirus particle with a single cork. (**E**) Cedratvirus particle with two corks. (**F**) Oval-shaped Pandoravirus particle, with its typical ostiole

host cell's cytoplasm, an electron-luminous viral factory is noticeable, accompanied by the presence of electron-dense amorphous structures, indicative of active viral replication and assembly processes (Figs. 2, 3D and E and 4B).

−3 yaraviruses: With icosahedral symmetry, the viral particles exhibit a diameter of approximately 80 nm. Within the host cell's cytoplasm, the viral factory has two different areas: a granular, containing replicated genomic units; and an electron-lucent, containing empty capsids (Figs. 2, 3C and 4E).

Considering the identification of isolates by PCR, the viruses underwent reactions corresponding to all available targets, including Pandoravirus, mimivirus, Marseillevirus, cedratvirus, pithovirus, and Yaravirus. PCR results corresponded with TEM identification for all isolates, comprising 1 Pandoravirus, 13 mimiviruses, 26 marseilleviruses, 24 cedratviruses, and 3 yaraviruses. Therefore, the PCR indicated that all the isolates identified (Sup. Table 3) by TEM as pitho/cedrat-like viruses were, actually, cedratviruses.

Considering the increasing prevalence of giant virus prospective research in laboratories worldwide, we deem it pertinent and valuable to examine the relationships among variables such as substrates, biomes, and isolates. However, it is important to note that this analysis is descriptive in nature. As previously mentioned, the representativeness of explored substrates and biomes may not be equivalent due to logistical and financial constraints.

New isolates were obtained from all types of samples tested (Fig. 5A). The largest quantity came from freshwater samples (45 isolates), which also exhibited the highest richness of isolated viral groups (at least 4 groups). From the less representative samples (soil and mud), only 1 Marseillevirus was isolated from soil samples, while 3 yaraviruses were obtained from mud samples. Saltwater and sewage samples yielded an equal number of isolates (9 isolates), but sewage samples showed greater richness of viral groups (3 groups) compared to saltwater samples (1 group).

The largest quantity of isolates was obtained from the Cerrado biome (30 new isolates), followed by the Atlantic Forest (24 isolates). However, the number of viral groups was the same for both (4 groups). From the Cerrado biome, isolates of Marseillevirus, mimivirus, Pandoravirus, and cedratviruses were obtained, while from the Atlantic Forest biome, isolates of mimivirus, Marseillevirus, cedratvirus, and Yaravirus were obtained (Fig. 5B).

In addition, representative viruses from each group were selected for genome sequencing to confirm their prior identification. For mimivirus, Pandoravirus, cedratvirus, and Marseillevirus, a phylogenetic tree was constructed using DNA polymerase, a common marker for giant virus phylogenetic analysis. In the case of Yaravirus, the major capsid protein (MCP) gene was utilized for tree

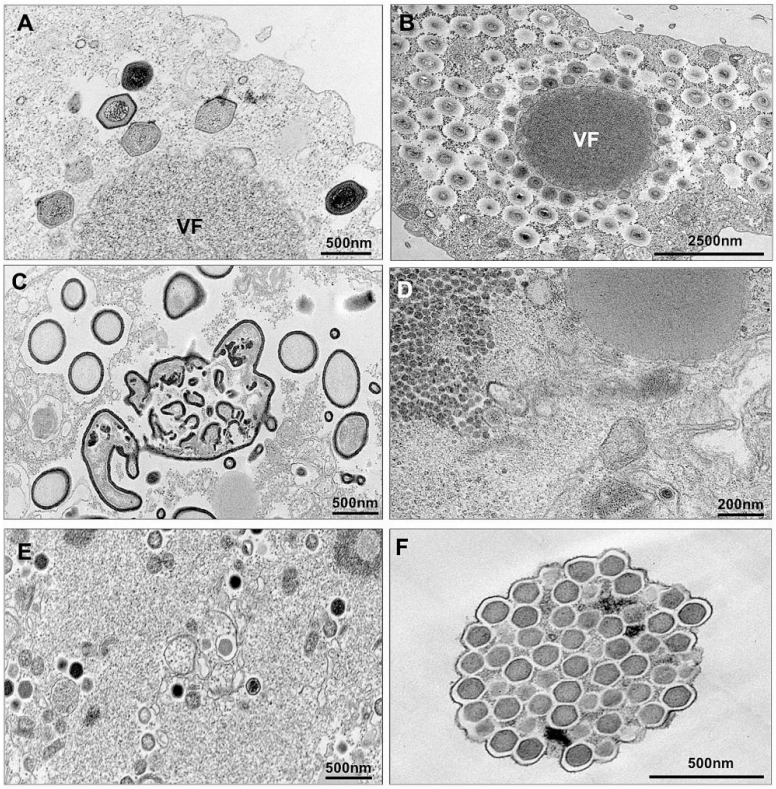

Fig. 4 Transmission electron microscopy of ***Acanthamoeba castellanii*cells infected by different viral isolates. (A and B)** Mimivirus spherical electron-dense viral factory with particles under morphogenesis. VF – viral factory. (**C**) Cedratvirus viral factory. In the center, an amorphous structure that likely is related to the tegument morphogenesis. (**D**) Yaravirus viral factory. At top-left, genomic granular area. It is possible to visualize a particle at bottom-right (**E**) Marseillevirus electron-lucent viral factory. It is possible to visualize viral particles in different stages of morphogenesis. (**F**) A Marseillevirus giant vesicle containing dozens of viral particles

construction. The results indicate that the previous identification of the isolates through TEM and PCR aligns with phylogenetic analysis, confirming the identification of the viruses (Fig. 6).

Discussion

The remarkable biodiversity of organisms in Brazilian natural and urban environments pose questions and supports virology studies employing a prospecting approach as employed in this study. Consequently, the ongoing isolation and characterization of giant viruses may contribute to the understanding of their ecological roles, evolutionary dynamics, and potential impact on host populations and ecosystems. In this sense, obtaining several environmental samples from different biomes, collected over a span of 11 years yielded a total of 67 new isolates belonging to at least five distinct viral groups, including Pandoravirus, mimivirus, Marseillevirus, cedratvirus, and Yaravirus. Over time, other prospecting studies on amoebal viruses were also performed in Brazilian biomes. Dornas et al. (2015) and Andrade et al. (2019) described the isolation of different types of *Acanthamoeba*-infecting giant viruses. However, in contrast to the present study, mimivirus was previously the most frequently isolated virus.

Information on the distribution patterns of isolated viruses across different biomes and sample types may be useful for future prospecting studies. Despite logistical and financial constraints, our study successfully captured the diversity of giant virus diversity present in Brazilian environments in a span of 11 years of research. However, it is important to acknowledge the limitations of our sampling approach, particularly regarding the representativeness of explored substrates and biomes. Future studies should aim to address these limitations and further

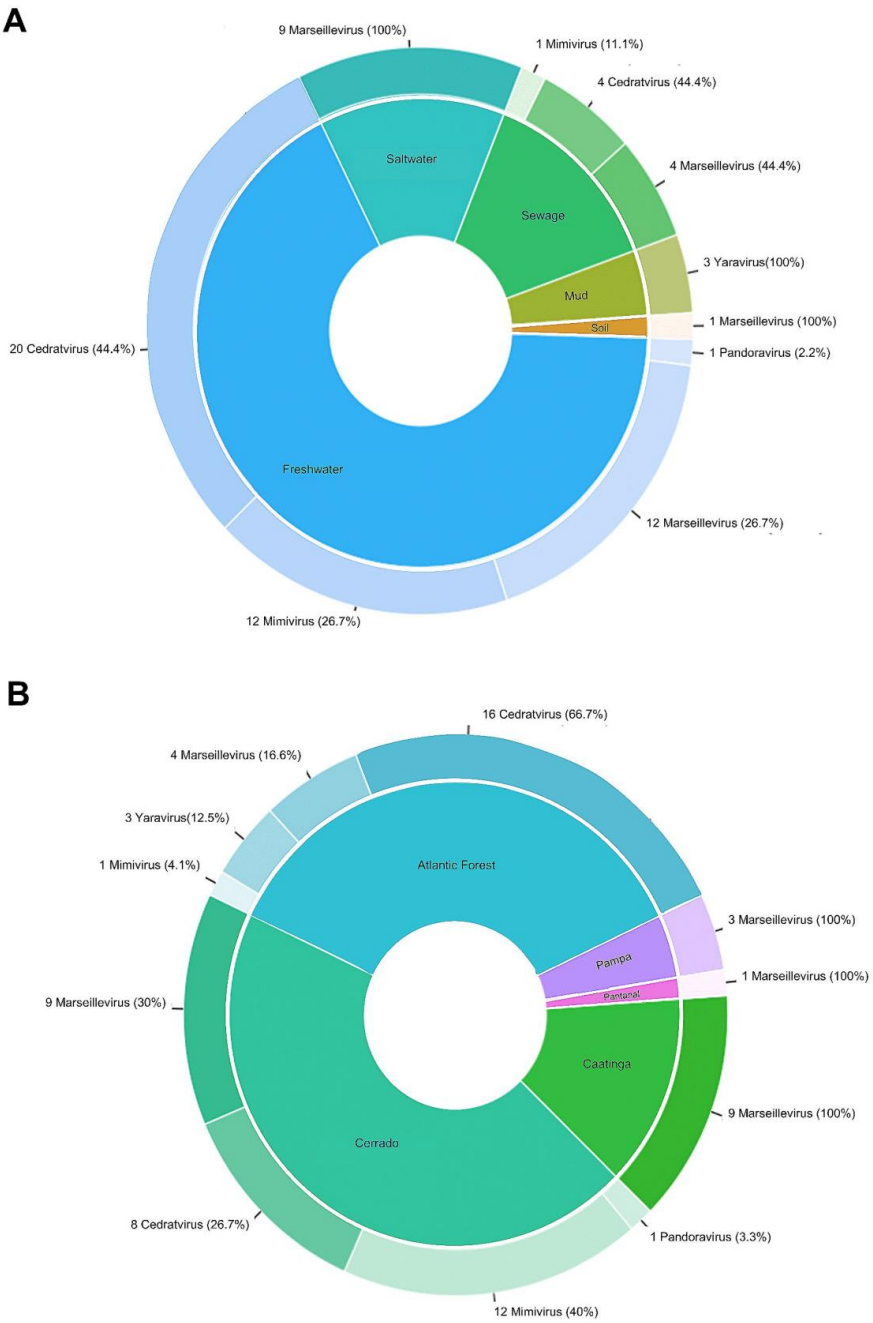

Fig. 5 Isolates, samples and biomes. (**A**) Quantity and representativeness of each viral group isolated within each type of sample (substrate). The greatest variety of viral groups was obtained in freshwater. (**B**) Quantity and representativeness of each viral group isolated within each biome. The greatest variety of viral groups was obtained in Cerrado samples

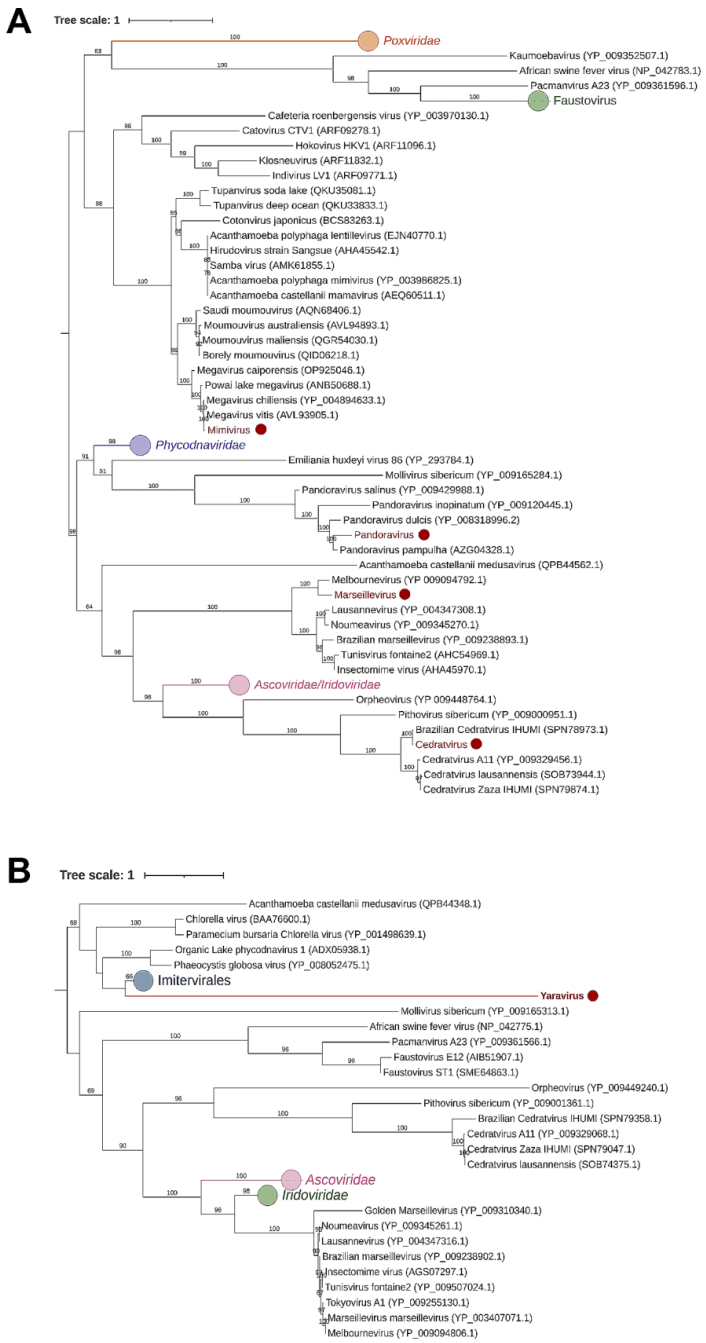

Fig. 6 Maximum likelihood phylogenetic trees constructed with amino acid sequences from the DNA polymerase subunit B (mimivirus, Pandoravirus, Marseillevirus and cedratvirus) and the major capsid protein (Yaravirus). The new isolates described here are red highlighted. The scale bar indicates the genetic distance

explore the ecological factors driving the distribution and diversity of giant viruses in Brazil and beyond.

Importantly, to an extended overview of such remarkable viral entities, our results suggest that morphological identification by TEM was effective in distinguishing between different viral groups. However, distinguishing between cedratviruses and pithoviruses proved challenging due to their strikingly similar characteristics. Therefore, additional PCR analysis or genome sequencing proved necessary to resolve identification ambiguities for these viruses. Last, phylogenetic analysis of representative viruses from each group further validated their taxonomic classification and evolutionary relationships. In conclusion, by combining such approaches we have expanded our understanding of the distribution of the giant viruses in Brazilian biomes. It is important to mention that some of the viral isolates described here have already undergone genomic and biological characterization [15, 21, 22]. This underscores the significance of prospecting studies as the initial step in knowledge construction.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12985-024-02404-z.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We thank the Virus Laboratory of the Federal University of Minas Gerais for all the support provided. We also thank the UFMG Microscopy Center, especially the technicians Denilson Cunha, Rodrigo Ferreira, Altair Mendes, Thalita Arantes, Marilene Oliveira and Breno Moreira, who collaborated from the preparation to the sample observation session. We are also grateful to each of the people who contributed to this work by donating samples: Adriana, Bet?nia, Edney, Natalia, Paula, Paulo, Poliana, Ivan, Isabela, Juliana, Severino, Silvana, Thalita. The text was revised using artificial intelligence (OpenAI - ChatGPT). All authors approved the final manuscript.

Author contributions

JSA and ACSPA designed and supervised the work. LSU and JPAJ sequenced the viral genomes. TBM, ILMA, BLA, MSS, MGB, performed experiments. AOC provided cells and assist on quality control. TBM, JSA, ER and LHR collected environmental samples. All authors reviewed the manuscript.

Funding

We would like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and Pró-Reitorias de Pesquisa e Pós-Graduação da UFMG (PRPG-UFMG) for the financial support. JPAJ, LHR and JSA are CNPq researchers.

Data availability

Viral sequences are available at Genbank under the accession numbers MT293574, OR343515, OR991738, MK131393.

Declarations

Ethical approval

The collections were carried out with authorization from the Biodiversity Authorization and Information System (SISBIO) under the identification numbers: 33326-2 and 34293-2; and SISGEN: A2291C9.

Sequencing data

Viral sequences are available at Genbank under the accession numbers MT293574, OR343515, OR991738, MK131393.

Competing interests

The authors declare no competing interests.

Received: 14 May 2024 / Accepted: 3 June 2024
Published online: 10 June 2024

References

1. Scola BL, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. A giant virus in Amoebae. Science. 2003;299:2033–2033. https://doi.org/10.1126/science.1081867
2. Dornas FP, Khalil JYB, Pagnier I, Raoult D, Abrahão J, La Scola B. Isolation of New Brazilian giant viruses from environmental samples using a Panel of Protozoa. Front Microbiol. 2015;6. https://doi.org/10.3389/fmicb.2015.01086
3. Andrade AC, dos Arantes SP, Rodrigues TS, Machado RAL, Dornas TB, Landell FP, Furst MF, Borges C, Dutra LGA, Almeida LAL. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. Virol J. 2018;15. https://doi.org/10.1186/s12985-018-0930-x
4. Mihara T, Koyano H, Hingamp P, Grimsley N, Goto S, Ogata H. Taxon Richness of Megaviridae exceeds those of Bacteria and Archaea in the Ocean. Microbes Environ. 2018;33:162–71. https://doi.org/10.1264/jsme2.ME17203
5. Silva LKDS, Andrade ACDSP, Dornas FP, Rodrigues RAL, Arantes T, Kroon EG, Bonjardim CA, Abrahão JS. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. Sci Rep. 2018;8(1):4000. https://doi.org/10.1038/s41598-018-22398-3. PMID: 29507337; PMCID: PMC5838162.
6. Kerepesi C, Grolmusz V. Giant viruses of the Kutch Desert. Arch Virol. 2016;161:721–4. https://doi.org/10.1007/s00705-015-2720-8
7. Kerepesi C, Grolmusz V. The Giant Virus Finder discovers an abundance of Giant viruses in the Antarctic Dry Valleys. Arch Virol. 2017;162:1671–6. https://doi.org/10.1007/s00705-017-3286-4
8. Schulz F, Alteio L, Goudeau D, Ryan EM, Yu FB, Malmstrom RR, Blanchard J, Woyke T. Hidden diversity of Soil Giant viruses. Nat Commun. 2018;9:4881. https://doi.org/10.1038/s41467-018-07335-2
9. Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, Dornas FP, Trindade GS, Ferreira PP, Marques JT, Robert C, Raoult D, Kroon EG, La Scola B, Abrahão JS. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virol J. 2014. https://doi.org/10.1186/1743-422X-11-95. PMID: 24886672; PMCID: PMC4113263.
10. Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic J-M, Ramus C, Bruley C, Labadie K, etal, et al. In-Depth study of Mollivirus Sibericum, a New 30,000-y-Old giant virus infecting Acanthamoeba. Proc Natl Acad Sci. 2015;112:E5327–35. https://doi.org/10.1073/pnas.1510795112
11. Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known Virosphere. Nat Commun. 2018;9. https://doi.org/10.1038/s41467-018-03168-1
12. Yoshikawa G, Blanc-Mathieu R, Song C, Kayama Y, Mochizuki T, Murata K, Ogata H, Takemura M. Medusavirus, a Novel large DNA virus discovered from Hot Spring Water. J Virol. 2019;93. https://doi.org/10.1128/jvi.02130-18
13. Boratto PVM, Serafim MSM, Witt ASA, Crispim APC, Azevedo BL. A brief history of giant viruses' studies in Brazilian biomes. Viruses. 2022;14:191. https://doi.org/10.3390/v14020191. MachadoT.B.deQueiroz, V.F.; Rodrigues, R.A.L.; et al.
14. Azevedo ALM, dos S. April. IBGE - Educa | Jovens Available online: https://educa.ibge.gov.br/jovens/conheca-o-brasil/territorio/18307-biomas-brasileiros.html (accessed on 11 2024).
15. Boratto PVM, Oliveira GP, Machado TB, Andrade ACSP, Baudoin J-P, Klose T, Schulz F, Azza S, Decloquement P, Chabrière E, et al. Yaravirus: a novel 80-Nm virus infecting Acanthamoeba Castellanii. Proc Natl Acad Sci. 2020;117:16579–86. https://doi.org/10.1073/pnas.2001637117

16. Machado TB, de Aquino ILM, Abrahão JS. Isolation of Giant viruses of Acanthamoeba Castellanii. Curr Protoc. 2022;2:e455. https://doi.org/10.1002/cpz1.455
17. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a New Genome Assembly Algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021
18. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015
19. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW, GenBank. Nucleic Acids Res. 2013;41:D36–42. https://doi.org/10.1093/nar/gks1195
20. Edgar RC, MUSCLE. Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. https://doi.org/10.1093/nar/gkh340
21. Machado TB, Picorelli ACR, de Azevedo BL, de Aquino ILM, Queiroz VF, Rodrigues RAL, Araújo JP Jr, Ullmann LS, Dos Santos TM, Marques RE, Guimarães SL, Andrade ACSP, Gularte JS, Demoliner M, Filippi M, Pereira VMAG, Spilki FR, Krupovic M, Aylward FO, Del-Bem L-E, Abrahão JS. Gene duplication as a major force driving the genome expansion in some giant viruses. J Virol. 2023 Dec 21;97(12):e0130923. https://doi.org/10.1128/jvi.01309-23
22. Pereira Andrade ACDS, Victor de Miranda Boratto P, Rodrigues RAL, Bastos TM, Azevedo BL, Dornas FP, Oliveira DB, Drumond BP, Kroon EG, Abrahão JS. New isolates of pandoraviruses: Contribution to the study of replication cycle steps. J Virol 2019 Feb 19;93(5):e01942–18. https://doi.org/10.1128/JVI.01942-18

Publisher's Note

viruses

MDPI

Review

A Brief History of Giant Viruses' Studies in Brazilian Biomes

Paulo Victor M. Boratto [1], Mateus Sá M. Serafim [1], Amanda Stéphanie A. Witt [1], Ana Paula C. Crispim [1], Bruna Luiza de Azevedo [1], Gabriel Augusto P. de Souza [1], Isabella Luiza M. de Aquino [1], Talita B. Machado [1], Victória F. Queiroz [1], Rodrigo A. L. Rodrigues [1], Ivan Bergier [2], Juliana Reis Cortines [3], Savio Torres de Farias [4], Raíssa Nunes dos Santos [5], Fabrício Souza Campos [5], Ana Cláudia Franco [5] and Jônatas S. Abrahão [1,*]

[1] Laboratório de Vírus, Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Minas Gerais, Brazil; pvboratto@gmail.com (P.V.M.B.); mateusmserafim@gmail.com (M.S.M.S.); asawitt1997@gmail.com (A.S.A.W.); anapbio2@gmail.com (A.P.C.C.); azvdobruna@gmail.com (B.L.d.A.); neogaps@gmail.com (G.A.P.d.S.); isabellaaquino92@gmail.com (I.L.M.d.A.); bastostalita04@gmail.com (T.B.M.); victoriafq18@gmail.com (V.F.Q.); rodriguesral07@gmail.com (R.A.L.R.)
[2] Embrapa Pantanal, Corumbá 79320-900, Mato Grosso do Sul, Brazil; bergiercpap@gmail.com
[3] Departamento de Virologia, Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-590, Rio de Janeiro, Brazil; cortines@micro.ufrj.br
[4] Laboratório de Genética Evolutiva Paulo Leminsk, Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa 58050-085, Paraíba, Brazil; stfarias@yahoo.com.br
[5] Laboratório de Virologia, Departamento de Microbiologia, Imunologia e Parasitologia, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre 90.050-170, Rio Grande do Sul, Brazil; engraissanunes@gmail.com (R.N.d.S.); camposvet@gmail.com (F.S.C.); anafranco.ufrgs@gmail.com (A.C.F.)
* Correspondence: jonatas.abrahao@gmail.com

Abstract: Almost two decades after the isolation of the first amoebal giant viruses, indubitably the discovery of these entities has deeply affected the current scientific knowledge on the virosphere. Much has been uncovered since then: viruses can now acknowledge complex genomes and huge particle sizes, integrating remarkable evolutionary relationships that date as early as the emergence of life on the planet. This year, a decade has passed since the first studies on giant viruses in the Brazilian territory, and since then biomes of rare beauty and biodiversity (Amazon, Atlantic forest, Pantanal wetlands, Cerrado savannas) have been explored in the search for giant viruses. From those unique biomes, novel viral entities were found, revealing never before seen genomes and virion structures. To celebrate this, here we bring together the context, inspirations, and the major contributions of independent Brazilian research groups to summarize the accumulated knowledge about the diversity and the exceptionality of some of the giant viruses found in Brazil.

Keywords: amoebae viruses; Brazilian isolates; giant virus; NCLDV; virosphere; virus diversity

check for updates

Citation: Boratto, P.V.M.; Serafim, M.S.M.; Witt, A.S.A.; Crispim, A.P.C.; Azevedo, B.L.d.; Souza, G.A.P.d.; Aquino, I.L.M.d.; Machado, T.B.; Queiroz, V.F.; Rodrigues, R.A.L.; et al. A Brief History of Giant Viruses' Studies in Brazilian Biomes. *Viruses* **2022**, *14*, 191. https://doi.org/10.3390/v14020191

Academic Editor: K. Andrew White

Received: 18 November 2021
Accepted: 15 January 2022
Published: 19 January 2022

1. Introduction

The description and characterization of the first amoebal giant virus (GV) in 2003, Acanthamoeba polyphaga mimivirus (APMV), raised important questions regarding the limits of the virosphere. These first findings revealed viral particles of about 700 nm, non-filterable through 0.2 µm pore size filters [1]. Although not the first described nucleocytoplasmic large DNA virus (NCLDV) (phylum *Nucleocytoviricota*), which includes other families such as *Poxviridae* [2], the original discovered member of the family *Mimiviridae* motivated new interpretations of crucial features in an organism recognized as a virus, advancing both knowledge of, and perspectives on, the most abundant group of organisms on Earth [3].

Much of the subsequent work on GVs has been driven by curiosity and the possibility of isolating novel groups of amoebal viruses and finding intriguing new characteristics. For instance, in 2008 La Scola et al. had previously isolated a distinct strain of APMV, the

acanthamoeba castellanii mamavirus, together with the first ever described virophage, the Sputnik virus (SNV), both found in a water-cooling tower of a hospital in France [4]. Here, the interest in describing more GVs contributed by influencing the consolidation of "virophages" as new satellite-like viruses, which are dependent on the mimivirus factory for their replication by putatively hijacking some key features (e.g., the viral RNA polymerase) [5].

Similarly, in 2009, Boyer et al. reported the isolation of the marseillevirus, a novel GV for which analysis of its core genes suggested a previously uncharacterized family of NCLDV. In addition, by unveiling some of the main features of the genome's repertoire for this new GV, the authors have proposed amoebas as potential "melting pots" of microbial evolution, given the convenient intracellular environment for gene transfer among parasites, including complex genomes that could advent from different GVs' viral sources [6]. Some of the GVs' hosts (different amoeba genus, e.g., *Acanthamoeba*) are indeed considered ubiquitous, found in almost all latitudes [7,8], as well as in a wide-range of environments, including wastewater [9], terrestrial and (deep) marine water [7,10], thermal springs [11], permafrost [12], ventilation and air conditioning systems, and even in hospital settings [13,14]. Notwithstanding, GVs can be found in a large set of different native hosts or host-associated organisms, from other various species of amoebas [8,15] to filtering feeding organisms such as oysters [16]. Recently, metagenomics studies have also indicated that GVs are even more abundant in marine environments than prokaryotes, suggesting that these viruses may play a fundamental role in nature as biological control agents, regulating biogeochemical cycles, and potentially acting as evolutionary driving forces [17–19]. Ultimately, hijacking or utilizing cellular components and translational machinery may indicate a common origin, regarding information on life's evolution, and the presence of translation proteins may open new hypotheses about GVs' origin and phylogenetic relationships with other domains of life [20].

The broad-spectrum environmental profile of GVs made Brazil an interesting field to search and study these microorganisms, especially considering the wide range and diversity of environments and biological dispersion throughout biomes, settings and native habitats around the country, such as the Amazon forest and Cerrado savannas, as well as the Pantanal wetlands, including the soda lakes of Nhecolândia in the middle of the woods, which hide rich organic sediments [7,10,21–23]. These characteristics were reflected in the findings and discoveries of several GV isolates in the Brazilian territory, such as: (i) tupanvirus soda lake, isolated from the Brazilian Pantanal (Nhecolândia lakes) and tupanvirus deep ocean, isolated from sediment at 3000 m below the water surface line at 'Bacia de Campos', in Rio de Janeiro [10]; (ii) samba virus (SMBV), isolated from the Brazilian Amazon [22]; (iii) cedratvirus getuliensis, from sewage samples in Minas Gerais state [24]; (iv) niemeyer virus [25], faustovirus mariensis [26] and yaravirus [27], all of them isolated from the Pampulha urban lagoon; (v) a number of *Mimivirus* isolates [7,21]; and many others (Figure 1A,B), some of which we will further discuss in this review.

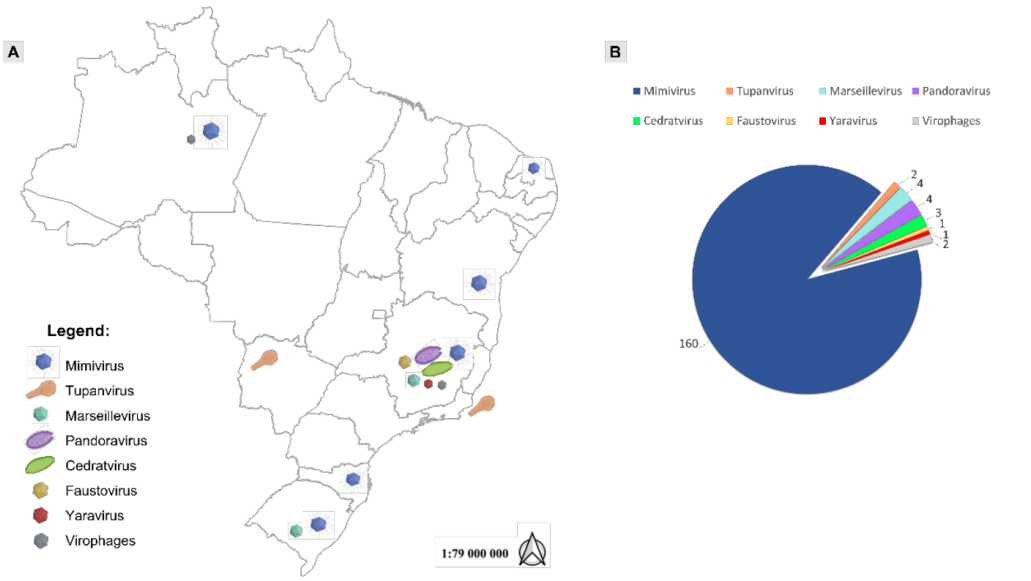

Figure 1. Location and numbers of giant viruses isolated in Brazil. (**A**) Schematic map showing the sites of isolation for the major groups of giant viruses discovered in Brazil. (**B**) Number of isolates discovered for each of these groups.

2. Giant Viruses Discovery and Isolation

2.1. Mimiviruses Boosted Amoebal Giant Viruses' Research

The first amoebal GV isolated in Brazil dates from 2011. During a field trip to the Brazilian Amazon, aiming to search for GVs, water samples were collected from the Negro River, in Manaus city. These samples were then assessed for prospection using *Acanthamoeba castellanii* cells as an isolation platform, which allowed the discovery of the first Brazilian GV, named samba virus (SMBV) [22].

Samba viruses have mimivirus-like particles, showing capsids with an average diameter of 527 nm, surrounded by fibrils of 155 nm [28]. The SMBV genome is composed of about 1.2 Mb, and the phylogenetic analysis clustered it within the lineage A of the mimiviruses (Table 1). Analysis of the SMBV replication cycle using a set of electron microscopy images showed several similarities with the APMV replication cycle. Moreover, these images also revealed the presence of smaller viral particles that were further confirmed as the first Brazilian virophage named Rio Negro virophage (RNV) [22]. A few years later, RNV genome was sequenced and assembled presenting 18,145 bp, very similar to the sputnik 2 virophage genome [29].

Later, a new sputnik-like virophage named guarani virophage was also isolated from water samples obtained in the Pampulha Lagoon, Belo Horizonte city, Minas Gerais state. A deep characterization of its genome (18,967 bp) was performed, and its replication is described as rather slow (replication at 4 h.p.i. and particles morphogenesis at 16 h.p.i.) when compared to the cycle of its associated GV [30].

After SMBV discovery, several mimiviruses belonging to the three currently known lineages (A, B and C) were isolated from different Brazilian environmental samples, such as the so-called "Br-mimiC", mimivirus golden (MVGD), isolated from golden mussels (*Limnoperna fortunei*) from Guaíba Lake, Rio Grande do Sul, in 2014 [31] and mimivirus gilmour (MVGM), isolated from water collected at the Pampulha Lagoon, in 2015 [21]. In this same work, the isolation of another 64 mimiviruses were described from water samples

collected at the Pampulha Lagoon. They were obtained from three different *Acanthamoeba* species (*A. castellanii*, *A. polyphaga* and *A. griffini*), and had representatives in the three lineages of mimiviruses [21]. Also in 2015, 20 mimiviruses belonging to the lineage A were obtained from oyster-related samples of three different coastal regions of Brazil [16]. Considering their water-filtering capacity, these bivalves were tagged as excellent sources for the isolation of new GVs because their body allows the accumulation of both viruses and amoebas [16].

Table 1. General features of Brazilian giant viruses with complete sequenced-genomes.

Group of Virus	Virus	Type of Sample	Location (Year of Isolation)	Genome Size (bp)	ORFs	ORFans	GC %	Reference
Mimiviridae (lineage A mimivirus)	Samba virus	Fresh water	Negro River (2011)	1,181,380	971	0	27	Campos et al., 2014
	Amazonia virus	Fresh water	Negro River (2011)	1,179,119	979	1 (0.1%)	27	Assis et al., 2015
	Kroon virus	Urban lake water	Lagoa Santa city (2012)	1,221,932	944	3 (0.3%)	27	Assis et al., 2015
	Oyster virus	Oysters	Santa Catarina state (2013)	1,200,220	948	1 (0.1%)	27	Assis et al., 2015
	Niemeyer virus	Urban lake water	Pampulha Lagoon (2011)	1,299,140	1003	0	28	Boratto et al., 2015
Mimiviridae (lineage B mimivirus)	Borely moumouvirus	Fresh Water	Serra do Cipó (2018)	1,038,187	947	3 (0.3%)	25.2	Silva et al., 2020
Mimiviridae (lineage C mimivirus)	Mimivirus gilmour	Urban lake water	Pampulha Lagoon (2014)	1,258,663	1135	28 (2.4%)	26	Assis et al., 2017
	Mimivirus golden	Golden mussels	Guaíba Lake (2014)	1,248,960	1127	19 (1.6%)	26	Assis et al., 2017
Mimiviridae	Tupanvirus deep ocean	Deep Ocean sediments	Campos dos Goytacazes city (2018)	1,439,508	1276	378 (29.6%)	28	Abrahão et al., 2018
	Tupanvirus soda lake	Soda Lake	Nhecolândia, Pantanal biome (2018)	1,516,267	1359	375 (27.6%)	28	Abrahão et al., 2018
Marseilleviridae	Brazilian marseillevirus	Sewage	Pampulha Lagoon (2014)	362,276	491	29 (5.9%)	43.3	Dornas et al., 2016
	Golden marseillevirus	Golden mussels	Guaíba Lake (2014)	360,610	483	43 (8.9%)	43.1	Santos et al., 2016
Cedratviruses	Brazilian cedratvirus	Water supplemented with biofloc	Belo Horizonte city (2018)	460,038	533	11 (2.1%)	42.9	Rodrigues et al., 2018
Faustovirus	Faustovirus mariensis	Urban lake water	Pampulha Lagoon (2019)	466,080	483	0	36	Borges et al., 2019
Yaravirus	Yaravirus brasiliensis	Muddy water	Pampulha Lagoon (2020)	44,924	74	68 (91.9%)	57.9	Boratto et al., 2020

In another study, a mimivirus-related isolate called Niemeyer virus (NYMV) was discovered, once again from water samples obtained from the Pampulha Lagoon [25]. NYMV has a genome of approximately 1,299,140 bp, harboring a set of duplicated aminoacyl-tRNA synthetases, which suggests that such duplications may be important for the evolutionary history of mimiviruses (Table 1). In 2017, another lineage A mimivirus was described, this time from water samples collected from an urban lake at the Lagoa Santa city, also in the Minas Gerais state, and it was named kroon virus (KV) (Figure 2A) [32]. The study of KV (1,221,932 bp genome [33]) has established an interesting view of the distinct ways by

which the major capsid protein (MCP) mRNA can be differentially processed, depending on the lineages of mimiviruses (Table 1) [32]. Apparently, for each of these mimiviruses' lineages there is a genetic layout concerning how the MCP gene is organized in terms of its exons/introns and how they are arranged. As an example, in the KV study the nucleotide sequences of the third exon of the MCP (observed in the genome of all mimi-viruses) was described to be an alternative marker to disentangle each of the three lineages. In addition, a different form of mature mRNA was also described in transcripts of the MCP for mimiviruses of a given lineage (e.g., APMV and KV) [32]. Subsequently, in 2018, 64 giant viruses of the *Mimiviridae* family (26 from lineage A, 13 from lineage B, two from lineage C and 23 from unidentified lineages) were described from various types of samples, including marine water from Antarctica, which was the first time to our knowledge that mimiviruses were isolated in this continent [7].

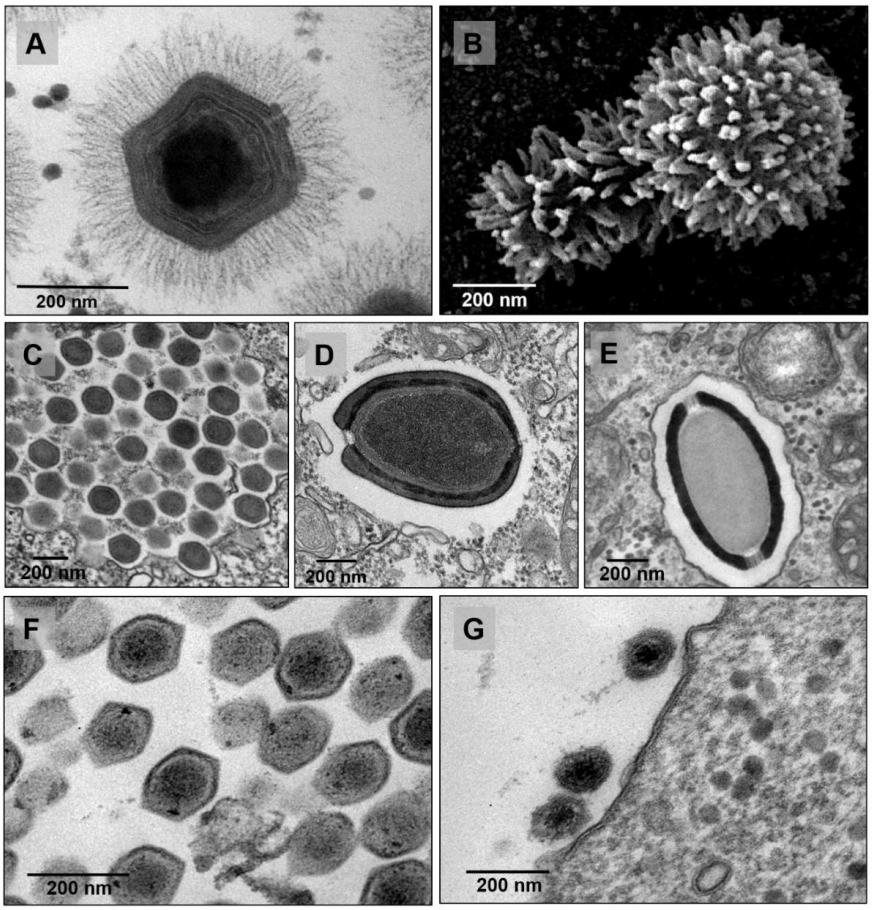

Figure 2. Panel with TEM images for the major groups of amoebal viruses isolated in Brazil. (**A**) mimivirus, (**B**) tupanvirus (source: 10.1038/s41598-018-36552-4), (**C**) marseillevirus, (**D**) pandoravirus, (**E**) cedratvirus, (**F**) faustovirus, and (**G**) yaravirus.

The year 2018 has also been marked by the description of one of the longest and most complex viruses described to date, obtained from a set of samples collected from

extreme aquatic environments. The tupanvirus soda lake was isolated from samples collected from an alkaline salty lake (Nhecolândia, Pantanal, Brazil) while the deep ocean tupanvirus was obtained from 3000 m depth sediment samples below the Atlantic Ocean at 'Campos dos Goytacazes' (Figure 2B) [10]. In contrast to other giant viruses previously isolated, these GVs are able to infect and establish a productive cycle in many species of amoebas, including *Acanthamoeba* spp., *Vermamoeba vermiformis*, *Dictyostelium discoideum* and *Willartia magna* [10]. The tupanviruses' capsid is similar to that of other mimiviruses already described in this review, around 450 nm in diameter, with a pseudo-icosahedral symmetry, covered by a layer of fibrils. They also present a "stargate" vertex, that is, a noticeable star-shaped opening at one capsid vertex. Interestingly, however, these viruses have a tail attached to the capsid, which is also covered with fibrils, a feature never seen before for amoebal viruses (Figure 2B). Due to the plasticity of this tail, the particles can vary from 1.2–2.3 µm in length, making it the longest virus ever described [10].

The genomes of the tupanviruses are complex and composed of double-stranded DNA of 1.44–1.51 Mb, encoding 1276–1425 predicted proteins (Table 1) [10]. Phylogenetic analysis using the DNA polymerase B family gene and other unique features exhibited by these viruses suggested that the tupanviruses group together with other mimiviruses form a distinct clade, which supported the proposal to form a new genus called "*Tupanvirus*" [10,34]. These viruses have been shown to be even more surprising, as deep genome analysis detected the largest translational apparatus ever described in the virosphere, with 20 aminoacyl-tRNA synthetases (aaRS), 67–70 tRNAs, in addition to other proteins in the translation process, such as translation factors (initiation, elongation and release) and proteins related to tRNA and mRNA maturation [10,34,35]. In addition, 20% of their genome is similar to genes originating from cellular organisms, with 9% from eukaryotes (of these, 3% originate from amoebas), 3% from archaea and 8% from bacteria [35].

These findings support data that demonstrate how other groups of organisms are relevant in studying the evolution of NCLDV genomes (Table 1). The fact that they have these genes shared with members of other cellular domains suggests that tupanviruses could also be found in non-extreme environments [35]. Altogether, the genetic arsenal of these and other mimiviruses within the virosphere add new levels of complexity to the understanding of the tree (or rhizome [36]) of life [20,37].

2.2. The Second Family Arises: The Discovery of Marseilleviruses

After the discovery of the first mimiviruses, the search for GVs intensified. In 2009, a virus named *Marseillevirus marseillevirus* was isolated in a biofilm from a water cooling tower in Paris, France [6], which gave rise to the family *Marseilleviridae*, officially recognized by the International Committee on Taxonomy of Viruses (ICTV) in 2013 (Figure 2C) [38]. Since then, other marseilleviruses have been isolated from different sources: (i) Lausannevirus, was discovered in water samples collected from the Seine river, in 2011 [39]; (ii) Cannes 8 virus was isolated from water in a cooling tower in Cannes, in 2013 [40]; (iii) tunisvirus and Fontaine Saint-Charles virus were isolated from freshwater collected in decorative fountains in Ariana, Tunisia, and in France, respectively [41,42]; (iv) insectomime virus was isolated from the internal organs and digestive tract of a dipteran drone fly's larvae [43]; (v) Senegalvirus was discovered during metagenomic analysis of the bacterial diversity in the human gut microbiota from a apparently healthy African individual, in 2012 [44,45]; (vi) In 2014, the genomic characterization of Melbournevirus was reported, isolated from a freshwater pond in Melbourne, Australia [46]; and (vii) Port-Miou virus, isolated from a sample from a brackish submarine spring, in the Cassis Port-Miou Calanque, France, in 2015 [47].

Furthermore, different phylogenetic lineages of marseillevirus have been described. Initially, the phylogenetic analysis suggested the existence of three distinct lineages: Lineage A, consisting of *Marseillevirus*, Cannes 8 virus, Senegalvirus and Melbournevirus; Lineage B, consisting solely of Lausannevirus; and Lineage C, consisting of tunisvirus and insectomime virus [42]. That was based on phylogenetic reconstructions carried out with core genes

including the NA polymerase B family, the VV A18 helicase, the D5 primase–helicase, the very late transcription factor 2B and the MCP [42].

However, the discovery of the first marseillevirus in America resulted in the creation of a new lineage in the family. The Brazilian Marseillevirus (BrMV) was described in 2014 from a sewage sample from a treatment station in the Pampulha lagoon [47]. The new lineage is supported by comparative genomic analyses highlighting several divergences between BrMV and other marseilleviruses (Figure 3) [47].

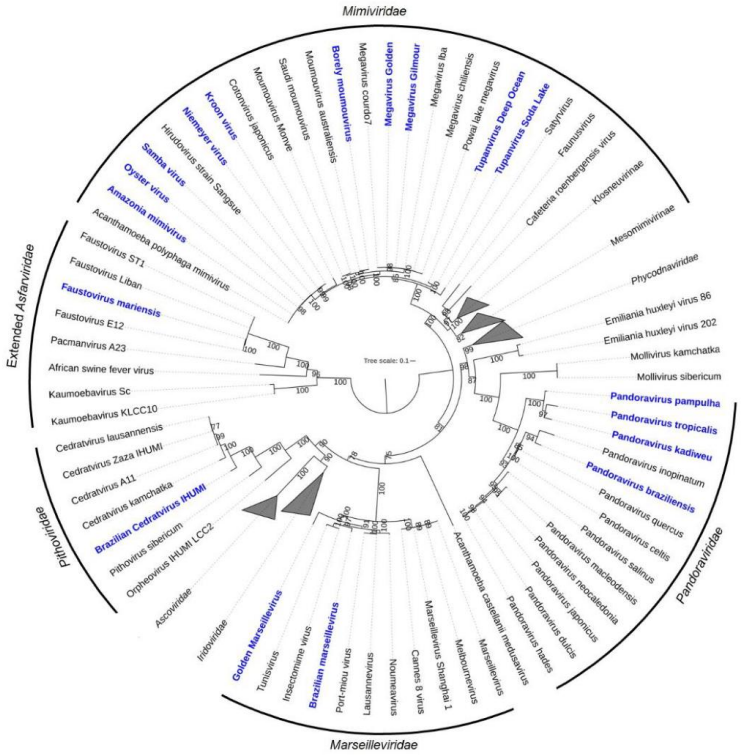

Figure 3. Maximum likelihood phylogenetic tree based on amino acid sequences of DNA polymerase B family of *Nucleocytoviricota*. Brazilian isolates are bold and highlighted in blue. Sequences were aligned using Muscle [48] and low conserved regions were removed using trimAl [49]. The tree was built using IQ-TREE [50] with 1000 ultrafast bootstrap replicates and the VT+F+R7 model chosen by ModelTest according to Bayesian Information Criterion. The tree was visualized in iToL [51]. The tree scale indicates the substitution rate.

A few years later, in 2016, the golden marseillevirus (GMar) was described as a new member isolated from golden mussels collected in southern Brazil [52]. The structure of the virus particles strongly resembled other marseilleviruses, with particles of approximately 200 nm, obtained from a coculture with A. polyphaga. The genome is composed of a circular dsDNA with 360,610 bp, comparable in size to the genomes of other members of the family *Marseilleviridae*, which range from 346,754 bp to 386,631 bp for lausannevirus and insectomime virus, respectively (Table 1) [52]. A total of 483 ORFs were characterized. Curiously, despite this genome size similarity, the GMar genes' content harbors 48.03% uncharacterized proteins. Many of these uncharacterized proteins can be considered as orphan genes (ORFans), also reported in other GVs such as Pandoravirus with

93% ORFans [53]. In addition, comparatively to the 212 genes shared among Brazilian marseillevirus, marseillevirus, lausannevirus, tunisvirus and golden marseillevirus, there are fourteen non-shared genes, of which seven are among the GMar genes [52].

2.3. Opening the GVs' Box: The Discovery of Pandoraviruses

Ten years after the discovery of the first GVs, the description of a brand-new group of amoebal viruses has led virologists to become again surprised, as a series of new paradigms started to be challenged and the study of modern virology advanced. At the time, the discovery and characterization of the pandoraviruses established for the first-time a group with viral particles with sizes as great as 1 µm in length and genomes that exceeded the mark of 2.5 Mb, with an astonishing number of 93% of genes without recognizable homologs in available databases (e.g., GenBank) (Figure 2D) [53].

Before the investigations in Brazil, it is important to mention the initial studies regarding the first representatives of this group. Starting from 2013, these discoveries were made: (i) *Pandoravirus salinus*, isolated from a superficial sediment layer collected at the mouth of the Tunquen river in Chile; and (ii) *Pandoravirus dulcis*, isolated from a mud taken at the bottom of a freshwater pond near Melbourne, Australia [53]. A couple of years later, a study led to the reinvestigation of an endosymbiont isolated from an *Acanthamoeba* strain and concluded, by whole genome sequencing, that this organism was in fact a pandoravirus isolate, named Pandoravirus inopinatum [54,55]. In another work, a newly characterized isolate called Pandoravirus celtis was used to investigate a putative scenario in which the genetic divergence among the different isolates of pandoraviruses was caused by an ability of these viruses to perform the creation of genes through a de novo microevolution process [56].

During the years 2018 to 2019, two different studies carried out a series of in-depth approaches focusing on establishing a detailed view of the diversity of pandoraviruses, their evolution processes and aspects of their replication cycle [57,58]. In the first study, three samples of pandoravirus were first isolated and named as pandoravirus quercus, pandoravirus neocaledonia and pandoravirus macleodensis. Their replication cycles were independently investigated and interestingly, for the first time, the mature particles of pandoraviruses were filmed while being exocytosed by vesicles which were full of viruses [57]. The genomes of these isolates were fully sequenced and a new stringent reannotation protocol was established. With this new methodology, the genetic analysis of different isolates suggested a still open pan-genome for GVs, in which each novel isolate is predicted to be responsible for contributing more than 50 additional genes [57].

For the second study, three novel Brazilian isolates were used: (i) pandoravirus kadiweu, coming from samples of water collected in the city of Bonito, Mato Grosso do Sul; (ii) pandoravirus pampulha, and (iii) pandoravirus tropicalis, both coming from samples of water from an artificial lake located at the city of Belo Horizonte, Minas Gerais [58]. Here, the microscopy analysis was an important tool, not only to reinforce some already established data but also to reveal new features of the virus replication. As for other GVs, within 30 min of infection the pandoravirus virions were phagocytosed and engulfed inside a host vesicle called the phagosome [57]. This structure quickly fuses with lysosome-like organelles and triggers the next stage of replication, which is the start of viral uncoating [53,57]. The next step involves an intense manipulation of the host cell and deep modification of the cytoplasm environment in order to make the region of viral morphogenesis, known as the viral factory. The loss of the cell nucleus and an intense recruitment of the host membranes and mitochondria are necessary for this. The beginning of viral morphogenesis does not seem to have a polarization, as thought earlier [53]. Finally, the viral cycle ends with the host cell lysis [53,57].

Interestingly, however, it was observed in some microscopy images that several pandoravirus particles were packaged inside vesicles and transported to the periphery of the host cell before amoebal disintegration. Additionally, one-step-growth curves have shown the beginning of viral release around 6 to 9 h post-infection, before the onset of

the amoebas' lysis [58]. These results, together with data that show a negative impact on pandoravirus release by cells treated with brefeldin (a membrane traffic inhibitor), suggest an important role of exocytosis for early liberation of pandoravirus particles in an amoeba infection [58]. Such observations are commonplace to other GVs with analogous replication cycles, including, for example, the cedratviruses described in the next section.

2.4. A Double-Corked GV: Isolation and Characterization of the Cedratviruses

Viruses belonging to the cedratvirus group were first detected in 2016, with the isolation of Cedratvirus A11, a viral representative coming from diverse environmental samples collected in Algeria [59]. Their structure is constructed by a ~1 µm ovoid-shaped particle, resembling some morphological features of the pithovirus virions, though with a notable difference: the presence of two corked regions (instead of a single one) at the extremities of the particle [59]. Their genome is composed of a circular dsDNA with about 590 kbp, and it has been found to share a close relationship with the genomes of the two currently known pithoviruses (both in size and in genome content), pithovirus sibericum and pithovirus massiliensis [59].

The second cedratvirus isolate, called cedratvirus lausannensis, was obtained in an attempt to look for amoebal-resisting bacteria inside a drinking water plant located at the Morsang-sur-Seine commune, in France [60]. Four other isolates have been discovered since then: (i) cedratvirus zaza IHUMI, deriving from samples of sterile distilled water collected near Toulon city, in France; (ii) Brazilian cedratvirus IHUMI, collected from water samples supplemented with bio-floc in Belo Horizonte city, Brazil; (iii) cedratvirus Kamchatka, obtained from a muddy grit soil collected next to a volcano area in Russia; and (iv) cedratvirus getuliensis (Figure 2E), collected from sewage samples from the Itaúna city, Brazil [24,61,62]. Interestingly, the isolate Brazilian cedratvirus IHUMI is a representative of the group which harbors both particle and genome sizes with remarkable differences in comparison with the other cedratviruses discovered so far. The virion is approximately 910 nm in length, with some of the particles reaching around 696 nm, and the genome is also smaller, with a DNA molecule of 460,038 bp (Table 1) [63]. Comparative genomic analysis also indicated that this Brazilian isolate is the founding member of a new lineage of cedratviruses (Figure 3) [63].

In 2018, through the analysis of a series of electron microscopy images and by performing biological assays, an interesting study has helped to reveal most of the steps in the replication of cedratviruses. As expected for an amoebal virus with large-sized virions, the viral cycle starts by the particles getting into the infected cell through the exploration of a phagocytic pathway that is physiologically presented by the host [24]. Corroborating this observation, lower titers of the cedratvirus virions are observed when the infected amoebal cells are pre-treated with cytochalasin D, an inhibitor of phagocytosis. The cycle then progresses to the formation of an electron-lucent viral factory (as large as the cellular nucleus) in the cytoplasm of the infected amoeba and, differently from observed during infection of most giant viruses, the cellular nucleus seems to remain intact [24].

However, some typical cellular alterations are still observed, such as the recruitment of mitochondria around the viral factory region, the polarization of lysosomal vesicles in the infected cell and an intense traffic of membranes which were seen to be important during the morphogenesis of cedratvirus virions [24]. This step is described as very complex and relies on the formation of several membrane precursors (or crescents) which later assume the correct conformation of a mature viral particle. Finally, the viral cycle ends with the mature particles released via cell lysis or exocytosis [24]. Cedratviruses also present structural similarities and infection features to other GVs, such as the orpheoviruses, as discussed below.

2.5. Another Amoeba, Another Virus: Discovery and Characterization of Orpheovirus

By implementing amoebas of the *Vermamoeba vermiformis* species as a platform of isolation, new groups of viruses were discovered from different samples. Among them, an

Orpheovirus was isolated in Marseille, France, from samples of rat stool [62]. Nevertheless, Souza et al. observed that, differently to previous findings of viruses infecting amoeba, CPE caused by Orpheovirus could be split into an early stage (3 to 12 h.p.i.), when cells stretch into a branched fusiform shape, and a late stage (starting at 24 h.p.i.), when cells become rounded [64].

The in-depth characterization of the replication cycle demonstrates that it takes approximately 30 h to be completed. It is suggested that one or more particles of Orpheovirus, which are around 1.1 µm, are phagocytized by the host cell within 1 h.p.i. [62,64]. After entry, the particle's internal content is released when the membrane that surrounds the viral core fuses with the endosomal membrane through a structure called ostiole, located at the apex of the particle. Subsequently, the formation of the large electron lucent viral factory is observed, concomitantly with the recruitment of mitochondria and membranes [64]. Membrane recruitment and bleb formation also seems to be important for the viral factory formation and particle morphogenesis since they are affected by treatment with a membrane trafficking inhibitor at the middle stage of infection (8 h.p.i.), which is also observed for cedratviruses [24]. Similarly, as described for other viruses, the particle morphogenesis initiates with the formation of electron-dense semicircular structures, which are filled with their internal content until the formation of the complete closed particle [64].

The complete particle presents smaller fibrils, when compared to mimiviruses, and at least two layers between the fibril layer and the inner membrane [62,64]. Finally, the infectious particles start to be released by exocytosis, detected in the supernatant at 12 h.p.i. Moreover, it is observed that cell lysis also plays a role in viral particle release, mostly at late timepoints of infection. Along with the infectious particles, the formation of defective particles is also observed [64].

2.6. The Isolation and Characterization of Faustoviruses

The faustoviruses are a group of giant viruses first detected in 2015 from samples of sewage from different regions in France and in Dakar, Senegal [65]. In Brazil, the first representative of this group was isolated and described in 2019, from prospecting studies of water samples from the Pampulha lagoon. Faustovirus mariensis, as it was called, is a virus with icosahedral particles reaching approximately 190 nm in diameter and inducing cytopathic effects on amoebas of the *Vermamoeba vermiformis* species (Figure 2F) [26]. Their genome is composed of a circular, double-stranded DNA molecule of about 466,080 bp (Table 1). Like other GVs, the f. mariensis replication cycle starts with the infection of the amoeba in its trophozoite form. This infection progresses to the formation of a large electron-lucent viral factory and the recruitment of mitochondria to its periphery [26]. The morphogenesis of f. mariensis is similar to that of other faustoviruses previously described in the literature, with new mature particles being formed in small honeycomb structures within the cytoplasm of the host cell. Lysis of the infected cell is the most important means of releasing the f. mariensis progeny described so far [26].

In a rare antiviral strategy described for GVs and their amoebal hosts, Borges et al. have observed that the infection of *Vermamoeba vermiformis* cultures is able to trigger a process of encystation of the neighboring cells, trapping the particles of f. mariensis inside their host and preventing further infection in the population of amoebas [26]. This event, considered to be observed for the first time in these viruses, was directly influenced by f. mariensis infection at a multiplicity of infection (MOI) dependent rates. When cysts were derived from cells infected at high MOIs, they were permanently incapable of excysting, therefore becoming trapped inside the particles of f. mariensis. However, when these amoebal cells came from infections at lower MOIs, only the cells with neither viral particles nor factories were able of excysting [26].

Faustoviruses are also phylogenetically related to kaumoebaviruses and asfarviruses, with the hypothesis that a common ancestor is shared between these viruses (Figure 3) [62]. After analysis considering this evolutionary proximity, motifs that play the role of promoter sequence in asfarvirus have been identified within the faustovirus genome, leading to the

conclusion that rich A-T (TATTT and TATATA) regions may also have an important role in the gene expression of both kaumoebavirus and faustovirus. These findings shed new light for a better understanding of giant virus's gene expression [66]. As aforementioned, intriguing information regarding the GVs' discovery and characteristics are quite common, and some unique factors have attracted attention in the field, such as the recent discovery of the yaravirus in 2020 [27].

2.7. Yaravirus, a Small Virus among the Giants

In late 2020, the discovery of a new lineage of dsDNA virus would enhance our knowledge on the diversity and evolution of viruses in amoeba. The yaravirus brasiliensis, as it was called, has been described as a novel virus of *Acanthamoeba castellanii,* harboring a genome of ~45 kbp enclosed in an icosahedral particle of about 80 nm in diameter [27]. Differently from any other virus isolated from acanthamoeba so far, this virus does not seem to share many of the features which are thought to represent the NCLDV, as it has neither a large particle nor a complex genome (Figure 4) [27]. This may indicate one of the following: (i) yaraviruses either belong to an extremely reduced group of amoebal viruses which are part of the NCLDVs; or (ii) these viruses represent the first discovered lineage of amoebal viruses that are not part of this complex group [27].

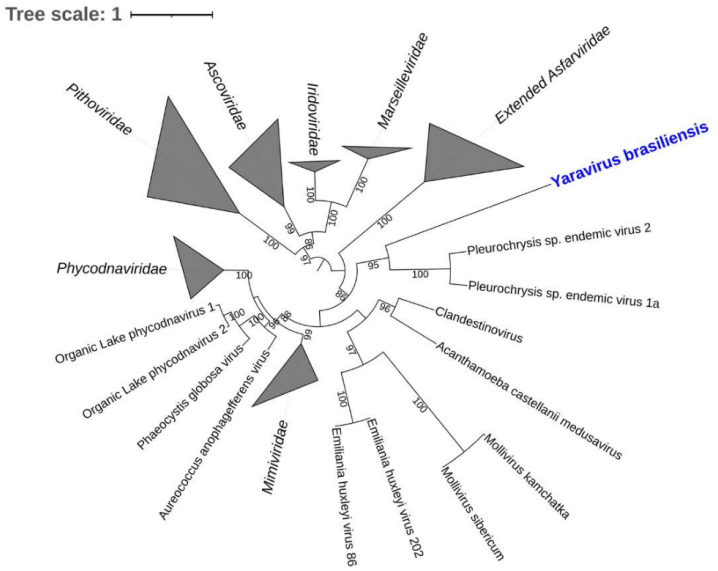

Figure 4. Maximum likelihood phylogenetic tree based on amino acid sequences of major capsid protein of Nucleocytoviricota. Yaravirus brasiliensis is bold and highlighted in blue. Sequences were aligned using Muscle [48] and low conserved regions were removed using trimAl [49]. The tree was built using IQ-TREE [50] with 1000 ultrafast bootstrap replicates and the VT+R3 model chosen by ModelTest according to Bayesian Information Criterion. The tree was visualized in iToL [51]. The tree scale indicates the substitution rate.

The genome of yaraviruses is composed mainly of ORFans, with an astonishing percentage of ~90% of their genes with functions never described before [27]. The search for yaravirus sequences in a huge dataset consisting of more than 8500 publicly available metagenomes from the most diverse habitats around the globe has also shown hits with distant homologs for the ATPase gene (NCVOG0249), with amino acid similarities that represented a number lower than 33% [27]. The discovery of yaravirus demonstrates

how important are studies focusing on isolating new viruses from the environment [27]. Although metagenomics analyses have an important role in describing new viral species by using their standard methods [67–69], microorganisms like yaraviruses would be very difficult to discover, as these protocols mostly involve recognition of genes already described [27]. This finding could also be seen as a marking point to revamp expositions in the virology field, from intriguing or persuading scientists to stimulating novel research and future researchers.

3. A Fight for Supremacy: Peculiar Features of GVs and Their Interaction with Amoeba Hosts

Much of the biology and particularities of mimivirus interactions with their hosts were discovered during early investigations of GVs. In this regard, both imaging and genomic techniques (e.g., electron microscopy, atomic force microscopy, sequencing, etc.) were of pivotal importance in uncovering many peculiar features of mimiviruses. APMV, for example, was observed to attach to the host cell through glycoside interactions between the long fibrils and surface glycans, with such adhesion also occurring with other unrelated organisms (e.g., arthropods and fungi), potentially facilitating the dispersion of these viruses in the environment [70].

Once they reach their hosts, differently from most non-giant viruses, mimiviruses (and most of the described GVs) enter the host cell through phagocytosis [71]. This was initially observed by transmission electron microscopy (TEM) analysis and further corroborated by biological assays, especially in cells treated with phagocytosis and endocytosis inhibitors [72]. At the apex of the mimivirus capsid there is a starfish-shaped protein complex that acts as a seal for the stargate, until the phagosome's internal environment promotes a new protein arrangement, unleashing the opening of the stargate and the release of the genome [73,74]. The acidification of the phagosome is suggested to be a factor that leads to capsid disassembly and membrane fusion [72,74,75].

In this regard, in 2011, the isolation of SMBV paved the way for further studies performed by other Brazilian research groups, enriching knowledge of mimiviruses' structure and biology. These studies included analysis of different Gs particles using distinct imaging techniques, such as cryo-electron microscopy (cryo-EM) and tomography, as well as fluorescence microscopy [28,76]. In a structural study developed by Schrad et al. (2020), for example, the viral particles of SMBV, tupanvirus, antarctica virus and mimivirus M4 were used to investigate the process of genome release in mimivirus-like particles [77]. Taking this work as an example, the authors have corroborated the importance of conditions such as temperature and pH for the opening of the vertex in these GVs. Here, new additional information on the viral uncoating was settled, as liberation of the viral seed (extra membrane sac) and the complete release of the viral genome were both manifested by experiments using specific conditions of these GVs' replication cycle (e.g., pH = 2 and/or 100 °C) [73]. Even though these conditions are non-biological, the authors suggest that they mimic GVs' replication cycle steps. It may also suggest that during the replication cycle other factors may play a role in capsid opening. In the same study, during the steps of viral genomic release and by adopting different imaging techniques such as cryo-EM, cryo-electron tomography and scanning electron microscopy (SEM), the authors have observed the formation of pockets devoid of DNA within the nucleocapsids of these GVs. Likewise, the analyses led to the identification of a set of proteins released from capsids during the early stages of infection within this whole complex [73]. Looking into another study, with analyses involving SEM and TEM, the authors have observed that for different mimiviruses the density of the fibrils on the surface of their capsid was variable and that this could be acquired simultaneously to genome acquisition throughout the process of morphogenesis in the large viral factories [72]. These techniques were also important in revealing key aspects of the replication cycle of different giant viruses. As already mentioned above, an antiviral strategy was beautifully described in a TEM study showing faustovirus mariensis particles trapped inside the cysts of the *Vermamoeba vermiformis* host

(Figure 5A) [26]. An in-depth description of the replication cycle of orpheovirus and cedratvirus was also established, mostly by imaging methods. For orpheovirus, we started to understand that viral exocytosis was as important as the cell lysis to the final step of this giant virus cycle (Figure 5B) [64]. For cedratvirus getuliensis, the contribution of these techniques has helped to describe a unique and complex sequential organization of the viral particle morphogenesis, including different steps of the formation of horseshoe and rectangular compartments, the incorporation of the second cork and thickening of the capsid well, and finally the formation of the ovoid-shaped virion (Figure 5C) [24]. Considering some intriguing observed features after the mimiviruses' release, another interesting study was proposed by Oliveira et al. (2019) and observed an aggregation of released tupanvirus particles with uninfected amoeba, promoting viral dissemination by the formation of host cell bunches (Figure 5D) [78]. This study revealed that this amoebal-bunch formation is correlated with the mannose-binding protein (MBP) gene expression, either induced by tupanviruses or between amoebas, through interactions among their receptor, both factors that may be important for the optimization of this process [78]. However, when we talk about the genome of tupanviruses, we observe that a great number of their genes are not present in many other mimiviruses' genomes [10], which may still hide some important information about these GVs' cycle. In view of how life has evolved on Earth, the complexity of this genome has also recently been used as an argument to suggest that viruses come from an ancestral strategy of life. According to the authors, in the period comprising the First to the Last Universal Common Ancestor (FUCA to LUCA), an intermediate ancestral (Transitional-LUCA) may have been arisen as an undifferentiated subsystem resembling a virus-like structure, from which most of the currently known viruses came [79]. Besides, aside from the already mentioned unique structural tail, and the formation of bunches [78], tupanviruses exhibited for the first time a cytotoxic phenotype to non-host cells [10]. These intriguing aspects metaphorically resemble a constant fight for supremacy [80] and help unravel the evolutionary history of GVs.

In addition to these distinct characteristics, it is worth mentioning that, as expected, the host cell does not remain indifferent to mimivirus infection. The encystment process is understood as a mechanism used by *Acanthamoeba* populations to become protected against several kinds of stressful conditions, such as dehydration, lack of nutrients, UV light, and viral infections, including against mimiviruses [26,81,82]. As observed in a study developed by Boratto et al. (2015), mimivirus infection is hampered even if those amoebas are not yet morphologically encysted but had already received the stimulus to turn into their resistant form (Figure 5E) [81]. Nonetheless, if the stimulus to become a cyst is triggered before the infection, mimiviruses as APMV are able to evade this protective status of the *Acanthamoeba* cyst, by preventing the expression of an encystment-mediating subtilisin-like serine protease and thus proceeding with the infection (Figure 5E) [81]. These studies demonstrate how complex are processes involving GVs' replication cycle and what an intricate interaction these viruses have with their amoebal hosts.

In addition to isolation studies already mentioned above, other works developed in Brazil have contributed to add knowledge of genomics and important relationships between marseilleviruses and their host. One of these studies has helped to bring light to pivotal processes in the replication cycle of marseilleviruses, specifically related to the viral entry and release. As extensively described in the literature, phagocytosis is a general route used by most GVs to enter amoebal cells [1,6,53,80]. This process is triggered only after recognition of particles larger than 500 nm [83]. However, viruses with particle sizes between 200 to 250 nm, as is the case for marseilleviruses, do not have the minimum size required to trigger this process [84].

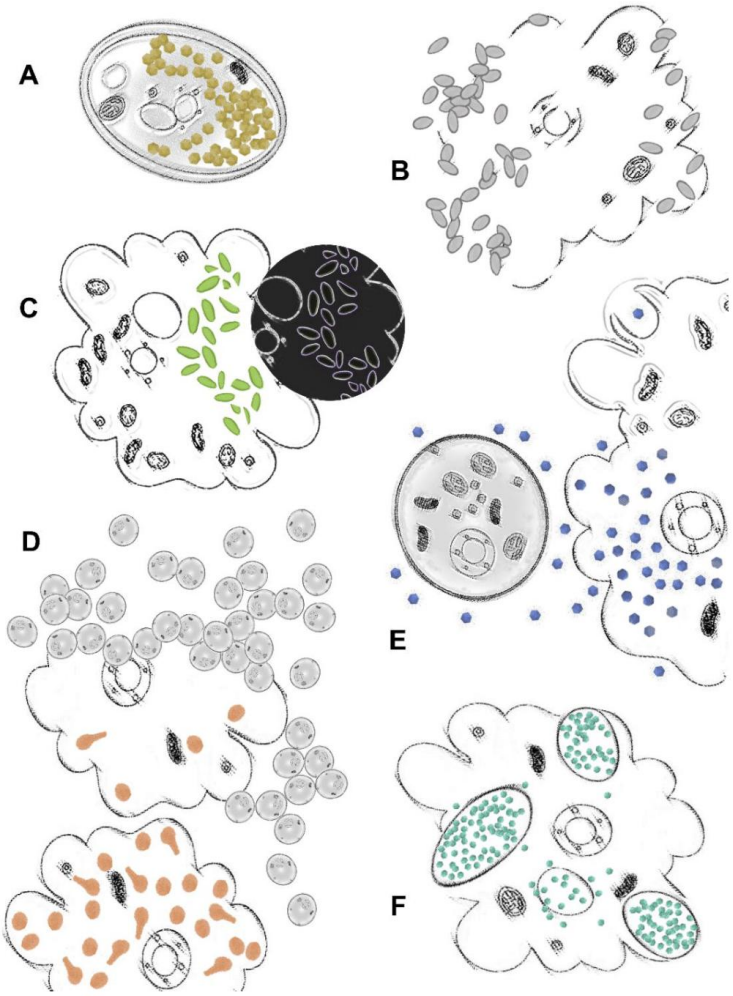

Figure 5. Unique features of giant viruses' (GVs') replication cycles unraveled in Brazil. (**A**) faustovirus dissemination is circumvented by amoebas with the enclosing of viral progeny inside the host's cysts [26]; (**B**) orpheovirus particles are released from the host by exocytosis or cell lysis [64]; (**C**) cedratvirus particles' morphogenesis follows a unique and complex sequential organization, including horseshoe and rectangular compartments, the incorporation of the second cork and thickening of the capsid well, and finally the formation of the ovoid-shaped virion [24]; (**D**) amoebas infected with tupanvirus are induced to aggregate with uninfected cells, forming giant host cell bunches [78]; (**E**) mimiviruses are able to infect amoebal trophozoites and prevent encystment, while cysts are resistant to infection [81]; (**F**) MsV are able to form giant vesicles with numerous viral particles derived from amoebal endoplasmic reticulum [84]. Amoeba images were generated from free vectors available online at Vecteezy: https://www.vecteezy.com (accessed on 22 September 2021).

By performing an in-depth investigation of the marseillevirus replication cycle and using a different set of virological assays (e.g., TEM, SEM, immunofluorescence, immunoblot-

ting), Arantes et al. have shown that during marseillevirus assembly the viral particles are organized inside large vesicles (some reaching about 3 µm in size) which are originated from the endoplasmic reticulum of the infected cells (Figure 5F) [84]. After viral release, those particles are then ready to infect another cell by exploring the phagocytosis of these vesicles that contain dozens to thousands of viral particles in their interior. In addition, viral release also seems to occur by individual virions. In this case, marseilleviruses exploit the endocytosis route to enter the cell by a mechanism which is dependent on acidification [84].

The *Marseilleviridae* family is also well known for its genomic mosaicism, which consists of the ability to incorporate foreign genes from other organisms that have *Acanthamoeba* as a common host [6]. Genomic studies of several strains of marseillevirus showed the presence of an A-T-rich promoter motif (AAATATTT) that is associated with 55% of the viral genes and that is conserved among all lineages. In addition, biological assays showed that the alteration of the promoter sequence negatively impacts the genes' transcription, showing a possible link of these sequences to the increased expression of some genes [85]. The presence of multiple copies of these motifs in the intergenic regions suggests that they may favor the fixation of newly acquired genes [85].

More recently, in 2020, analysis of the marseillevirus transcriptome revealed a temporal gene expression profile, indicating the existence of three categories: early, intermediate and late [86]. Genes belonging to different functional groups exhibited distinct expression levels throughout the infection cycle and marseillevirus infection causes significant changes in the host's transcription machinery, downregulating many genes [86].

Finally, it is worth mentioning that much of the features above described for GVs and their hosts have an influence directly affected by the intracellular environment of the amoebas. This environment has already been seen as an ecological site that comprehends a number of different and phylogenetic distant microorganisms, which not only inhabit the same location but are also observed to be in a strong process of coevolution. Even if not genetically related, an important portion of the genomic signatures (described as "the total net response to selective pressures") of the coevolving microorganisms are found to be incredibly conserved. This makes the intracellular environment of the amoebal host a sanctuary for interactions among several species of ecological and biomedical relevance [87].

4. Giant Viruses As a Tool to Update and Inspire: From the Research Fields to the Classroom

Since the known virosphere is notably anthropocentric, virology classes usually present viruses as pathogenic organisms, strongly associated with human diseases [3,88]. Instead of presenting these organisms as important tools of natural selection, ecological balance and the Earth's biogeochemical cycles, the commonly used material for teaching virology leads to a biased misconception of viruses as strictly bad, generating a certain fear in the students [89]. Besides this, other problems are the high cost of ensuring biosafety for practical virology classes, and motivation and mastery of the subject by the teachers (especially at the elementary school level) [90]. A further point is that viruses are typically very abstract for students, mainly due to their size, which limits their visualization to schematic figures, illustrations, and electron microscopy images [88].

The expansion of the perception of the virosphere by the giant viruses has unleashed a new way of understanding and teaching virology. Due to their colossal particles, the size limitation has been considered obsolete, turning these viruses into excellent learning tools [88,91]. Therefore, GVs can be visualized by optical microscopy, like bacteria and fungi, which are traditionally presented to students through common microscopes. Moreover, since they infect free-living amoebae, they represent a safe and low-cost instrument for practical virology classes [88,91].

In 2020, our group developed an educational kit to update the content typically taught in virology classes and align it to recent breakthroughs in virus research. Using slides, staining materials, viruses from the laboratory stock, and cell lineages, a microscope

slides kit called "Virus Goes Viral" was created (Figure 6A) [91]. This allows students to observe giant viruses' particles (Figure 6B–D), viral factories, and different lysis plaques in important viruses that infect animals [91]. As basic research regarding GVs in the Brazilian territory may enrich the knowledge of these microorganisms in the virology field, this kit may also aim to foster an inspiring learning environment, as well as ignite more interest in these fascinating organisms in the future.

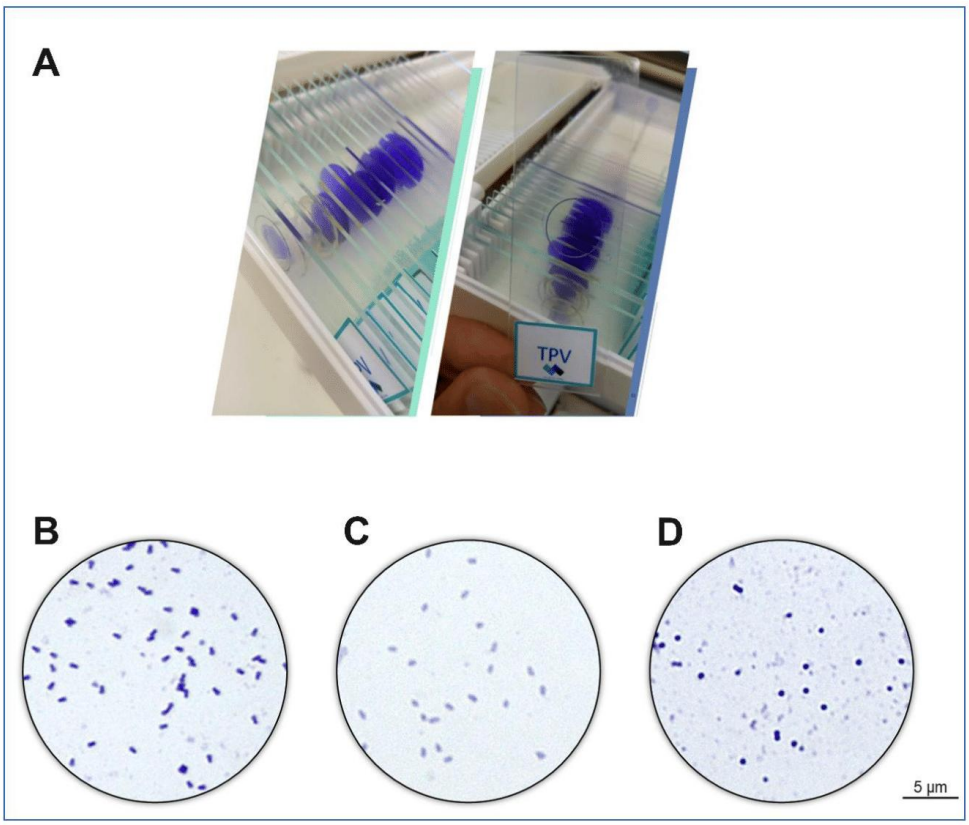

Figure 6. Optical microscopy images. (**A**) Source: Reference [91]. Visualization under 1000 times magnification of the stained purified particles of (**B**) tupanvirus, (**C**) cedratvirus and (**D**) Niemeyer virus, respectively.

5. Conclusions

The serendipitous discovery of APMV in 2003 changed the concept of viruses and expanded the limits of the virosphere [1]. Over the last two decades, different groups of large and giant viruses of amoebae have been described throughout the world, revealing many unusual particles' shapes and genome length and content. Culture-independent studies have proved the ubiquity and the astonishing diversity of giant viruses on Earth [18,68]. Now, we must go deeper into the characterization of these viruses, by using different isolates as models. Several new viruses isolated from distinct viral families (e.g., *Mimiviridae* and *Marseilleviridae*) are now under in-depth investigation to better understand their biology and evolution, and these outstanding discoveries may even change our perception of

life itself (e.g., *Mimiviridae* as a new branch derived from a population that gave origin to the modern Eukarya [20]).

Brazil is the fifth largest country in terms of territory and harbors different biomes, making the country an important global hub of tropical biodiversity. Over the past 10 years, the diversity of amoebal viruses in Brazilian environments has been uncovered, with hundreds of isolates belonging to distinct groups, including members of *Mimiviridae*, *Marseilleviridae*, Pandoraviridae, Pithoviridae, Faustoviridae, *Lavidaviridae*, among others, as discussed in previous sections. Interestingly, the most complex giant viruses described to date were isolated from two distinct areas in Brazil [10]. Furthermore, the recent discovery of the small mysterious yaravirus in the country highlighted the importance of continuing the search for new isolates, which could reveal completely new entities on Earth [27]. Brazilian groups, working alongside other experts in the field, have contributed to uncovering this unusual and exciting side of the virosphere. The study of amoebal viruses has already changed our perception of basic virology. Furthermore, giant viruses have recently been proposed as tools to improve virology learning at different educational levels [91]. Surely, other potential applications for these viruses are waiting to be revealed as new data emerges. This is an open field for remarkable discoveries, and we can expect great innovations as new amoebal viruses are isolated and characterized. In this context, the preservation of Brazilian biomes is a sine qua non condition, not only for the discovery of novel biological entities (including giant viruses), but also because of climatic, philosophical, political and economic reasons. Finally, we would like to reinforce that, although this review is a celebration of the 10th anniversary of giant virus studies in Brazilian biomes, we do not support any kind of excessive scientific nationalism. We are aware that studies of giant viruses in Brazil represent a modest contribution to the giant viruses' universe. This field has been constructed by remarkable worldwide research groups, and we are very grateful for their efforts and inspiring work.

Author Contributions: Conceptualization, P.V.M.B., M.S.M.S., A.S.A.W., A.P.C.C., B.L.d.A., G.A.P.d.S., I.L.M.d.A., T.B.M., V.F.Q., R.A.L.R. Writing—original draft preparation, P.V.M.B., M.S.M.S., A.S.A.W., A.P.C.C., B.L.d.A., G.A.P.d.S., I.L.M.d.A., T.B.M., V.F.Q., R.A.L.R. Writing—review and editing, P.V.M.B., M.S.M.S., I.B., J.R.C., S.T.d.F., R.N.d.S., F.S.C., A.C.F., J.S.A. Supervision and funding acquisition, J.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: JSA is a CNPq researcher (302081/2018-6).

Institutional Review Board Statement: The data here presented was partially or fully registered at SISGEN—numbers AA3B21E, A702EB8, A25764F, AC31840, A473BD3, A3DAB3F, AC3045D, A96431C, ABF23CC, A2F8816, A580BBD, AEC3EAA. Collection authorization -SISBIO numbers 33326, 34293 and 80252.

Informed Consent Statement: Not applicable.

Data Availability Statement: Genomic data can be found at genbank (https://www.ncbi.nlm.nih.gov/genbank/, accessed on 11 January 2022).

Acknowledgments: We would like to thank all colleagues that contributed to studies on giant viruses in Brazil and worldwide. It has been a long, hard, challenging but pleasant journey. We are grateful to Bernard La Scola (Aix Marseille Université, France) for those years of fruitful collaboration. We also thank the institutes that provided support in the last 10 years, including the CNPq, CAPES, FAPEMIG, Ministério da Saúde, Ministério de Meio Ambiente, Ministério da Educação and Pró-Reitoria de Pesquisa e de Pós-Graduacão da UFMG and Centro de Microscopia da UFMG.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Scola, B.L.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J.-M.; Raoult, D. A Giant Virus in Amoebae. *Science* **2003**, *299*, 2033. [CrossRef]
2. Iyer, L.M.; Aravind, L.; Koonin, E.V. Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *J. Virol.* **2001**, *75*, 11720–11734. [CrossRef] [PubMed]

3. Rodrigues, R.A.; Andrade, A.C.; Boratto, P.V.d.M.; Trindade, G.d.S.; Kroon, E.G.; Abrahão, J.S. An Anthropocentric View of the Virosphere-Host Relationship. *Front. Microbiol.* **2017**, *8*, 1673. [CrossRef] [PubMed]
4. La Scola, B.; Desnues, C.; Pagnier, I.; Robert, C.; Barrassi, L.; Fournous, G.; Merchat, M.; Suzan-Monti, M.; Forterre, P.; Koonin, E.; et al. The Virophage as a Unique Parasite of the Giant Mimivirus. *Nature* **2008**, *455*, 100–104. [CrossRef] [PubMed]
5. Koonin, E.V.; Yutin, N. Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology* **2010**, *53*, 284–292. [CrossRef]
6. Boyer, M.; Yutin, N.; Pagnier, I.; Barrassi, L.; Fournous, G.; Espinosa, L.; Robert, C.; Azza, S.; Sun, S.; Rossmann, M.G.; et al. Giant Marseillevirus Highlights the Role of Amoebae as a Melting Pot in Emergence of Chimeric Microorganisms. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21848–21853. [CrossRef] [PubMed]
7. Andrade, A.C.D.S.P.; Arantes, T.S.; Rodrigues, R.A.L.; Machado, T.B.; Dornas, F.P.; Landell, M.F.; Furst, C.; Borges, L.G.A.; Dutra, L.A.L.; Almeida, G.; et al. Ubiquitous Giants: A Plethora of Giant Viruses Found in Brazil and Antarctica. *Virol. J.* **2018**, *15*, 22. [CrossRef]
8. Aherfi, S.; Colson, P.; La Scola, B.; Raoult, D. Giant Viruses of Amoebas: An Update. *Front. Microbiol.* **2016**, *7*, 349. [CrossRef]
9. Schulz, F.; Yutin, N.; Ivanova, N.N.; Ortega, D.R.; Lee, T.K.; Vierheilig, J.; Daims, H.; Horn, M.; Wagner, M.; Jensen, G.J.; et al. Giant Viruses with an Expanded Complement of Translation System Components. *Science* **2017**, *356*, 82–85. [CrossRef]
10. Abrahão, J.; Silva, L.; Silva, L.S.; Khalil, J.Y.B.; Rodrigues, R.; Arantes, T.; Assis, F.; Boratto, P.; Andrade, M.; Kroon, E.G.; et al. Tailed Giant Tupanvirus Possesses the Most Complete Translational Apparatus of the Known Virosphere. *Nat. Commun.* **2018**, *9*, 749. [CrossRef] [PubMed]
11. Yoshikawa, G.; Blanc-Mathieu, R.; Song, C.; Kayama, Y.; Mochizuki, T.; Murata, K.; Ogata, H.; Takemura, M. Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. *J. Virol.* **2019**, *93*, e02130-18. [CrossRef] [PubMed]
12. Legendre, M.; Lartigue, A.; Bertaux, L.; Jeudy, S.; Bartoli, J.; Lescot, M.; Alempic, J.-M.; Ramus, C.; Bruley, C.; Labadie, K.; et al. In-Depth Study of Mollivirus Sibericum, a New 30,000-y-Old Giant Virus Infecting Acanthamoeba. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5327–E5335. [CrossRef] [PubMed]
13. Abrahão, J.S.; Dornas, F.P.; Silva, L.C.; Almeida, G.M.; Boratto, P.V.; Colson, P.; La Scola, B.; Kroon, E.G. Acanthamoeba Polyphaga Mimivirus and Other Giant Viruses: An Open Field to Outstanding Discoveries. *Virol. J.* **2014**, *11*, 120. [CrossRef] [PubMed]
14. Marciano-Cabral, F.; Cabral, G. *Acanthamoeba* spp. as Agents of Disease in Humans. *Clin. Microbiol. Rev.* **2003**, *16*, 273–307. [CrossRef] [PubMed]
15. Colson, P.; La Scola, B.; Raoult, D. Giant Viruses of Amoebae: A Journey through Innovative Research and Paradigm Changes. *Annu. Rev. Virol.* **2017**, *4*, 61–85. [CrossRef]
16. Andrade, K.R.; Boratto, P.P.V.M.; Rodrigues, F.P.; Silva, L.C.F.; Dornas, F.P.; Pilotto, M.R.; La Scola, B.; Almeida, G.M.F.; Kroon, E.G.; Abrahão, J.S. Oysters as Hot Spots for Mimivirus Isolation. *Arch. Virol.* **2015**, *160*, 477–482. [CrossRef]
17. Mihara, T.; Koyano, H.; Hingamp, P.; Grimsley, N.; Goto, S.; Ogata, H. Taxon Richness of "Megaviridae" Exceeds Those of Bacteria and Archaea in the Ocean. *Microbes Environ.* **2018**, *33*, 162–171. [CrossRef] [PubMed]
18. Moniruzzaman, M.; Martinez-Gutierrez, C.A.; Weinheimer, A.R.; Aylward, F.O. Dynamic Genome Evolution and Complex Virocell Metabolism of Globally-Distributed Giant Viruses. *Nat. Commun.* **2020**, *11*, 1710. [CrossRef]
19. Moniruzzaman, M.; Weinheimer, A.R.; Martinez-Gutierrez, C.A.; Aylward, F.O. Widespread Endogenization of Giant Viruses Shapes Genomes of Green Algae. *Nature* **2020**, *588*, 141–145. [CrossRef]
20. Marcelino, V.M.; Espinola, M.V.P.C.; Serrano-Solis, V.; Farias, S.T. Evolution of the Genus Mimivirus Based on Translation Protein Homology and Its Implication in the Tree of Life. *Genet. Mol. Res.* **2017**, *16*, 1–7. [CrossRef]
21. Dornas, F.P.; Khalil, J.Y.B.; Pagnier, I.; Raoult, D.; Abrahão, J.; La Scola, B. Isolation of New Brazilian Giant Viruses from Environmental Samples Using a Panel of Protozoa. *Front. Microbiol.* **2015**, *6*, 1086. [CrossRef] [PubMed]
22. Campos, R.K.; Boratto, P.V.; Assis, F.L.; Aguiar, E.R.; Silva, L.C.; Albarnaz, J.D.; Dornas, F.P.; Trindade, G.S.; Ferreira, P.P.; Marques, J.T.; et al. Samba Virus: A Novel Mimivirus from a Giant Rain Forest, the Brazilian Amazon. *Virol. J.* **2014**, *11*, 95. [CrossRef]
23. Guerreiro, R.L.; Bergier, I.; McGlue, M.M.; Warren, L.V.; de Abreu, U.G.P.; Abrahão, J.; Assine, M.L. The Soda Lakes of Nhecolândia: A Conservation Opportunity for the Pantanal Wetlands. *Perspect. Ecol. Conserv.* **2019**, *17*, 9–18. [CrossRef]
24. Dos Santos Silva, L.K.; Andrade, A.C.S.P.; Dornas, F.P.; Rodrigues, R.A.L.; Arantes, T.; Kroon, E.G.; Bonjardim, C.A.; Abrahão, J.S. Cedratvirus Getuliensis Replication Cycle: An in-Depth Morphological Analysis. *Sci. Rep.* **2018**, *8*, 4000. [CrossRef]
25. Boratto, P.V.M.; Arantes, T.S.; Silva, L.C.F.; Assis, F.L.; Kroon, E.G.; La Scola, B.; Abrahão, J.S. Niemeyer Virus: A New Mimivirus Group A Isolate Harboring a Set of Duplicated Aminoacyl-TRNA Synthetase Genes. *Front. Microbiol.* **2015**, *6*, 1256. [CrossRef]
26. Borges, I.; Rodrigues, R.A.L.; Dornas, F.P.; Almeida, G.; Aquino, I.; Bonjardim, C.A.; Kroon, E.G.; La Scola, B.; Abrahão, J.S. Trapping the Enemy: *Vermamoeba vermiformis* Circumvents Faustovirus Mariensis Dissemination by Enclosing Viral Progeny inside Cysts. *J. Virol.* **2019**, *93*, e00312-19. [CrossRef]
27. Boratto, P.V.M.; Oliveira, G.P.; Machado, T.B.; Andrade, A.C.S.P.; Baudoin, J.-P.; Klose, T.; Schulz, F.; Azza, S.; Decloquement, P.; Chabrière, E.; et al. Yaravirus: A Novel 80-Nm Virus Infecting Acanthamoeba Castellanii. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 16579–16586. [CrossRef]
28. Schrad, J.R.; Young, E.J.; Abrahão, J.S.; Cortines, J.R.; Parent, K.N. Microscopic Characterization of the Brazilian Giant Samba Virus. *Viruses* **2017**, *9*, 30. [CrossRef] [PubMed]

29. Borges, I.A.; de Assis, F.L.; Silva, L.K.D.S.; Abrahão, J. Rio Negro Virophage: Sequencing of the near Complete Genome and Transmission Electron Microscopy of Viral Factories and Particles. *Braz. J. Microbiol.* **2018**, *49* (Suppl. S1), 260–261. [CrossRef] [PubMed]
30. Mougari, S.; Bekliz, M.; Abrahao, J.; Di Pinto, F.; Levasseur, A.; La Scola, B. Guarani Virophage, a New Sputnik-Like Isolate From a Brazilian Lake. *Front. Microbiol.* **2019**, *10*, 1003. [CrossRef]
31. Assis, F.L.; Franco-Luiz, A.P.M.; Dos Santos, R.N.; Campos, F.S.; Dornas, F.P.; Boratto, P.V.M.; Franco, A.C.; Abrahao, J.S.; Colson, P.; Scola, B.L. Genome Characterization of the First Mimiviruses of Lineage C Isolated in Brazil. *Front. Microbiol.* **2017**, *8*, 2562. [CrossRef] [PubMed]
32. Boratto, P.V.M.; Dornas, F.P.; da Silva, L.C.F.; Rodrigues, R.A.L.; Oliveira, G.P.; Cortines, J.R.; Drumond, B.P.; Abrahão, J.S. Analyses of the Kroon Virus Major Capsid Gene and Its Transcript Highlight a Distinct Pattern of Gene Evolution and Splicing among Mimiviruses. *J. Virol.* **2018**, *92*, e01782-17. [CrossRef]
33. Assis, F.L.; Bajrai, L.; Abrahao, J.S.; Kroon, E.G.; Dornas, F.P.; Andrade, K.R.; Boratto, P.V.M.; Pilotto, M.R.; Robert, C.; Benamar, S.; et al. Pan-Genome Analysis of Brazilian Lineage A Amoebal Mimiviruses. *Viruses* **2015**, *7*, 3483–3499. [CrossRef] [PubMed]
34. Rodrigues, R.A.L.; Mougari, S.; Colson, P.; La Scola, B.; Abrahão, J.S. "Tupanvirus", a New Genus in the Family Mimiviridae. *Arch. Virol.* **2019**, *164*, 325–331. [CrossRef]
35. De Miranda Boratto, P.V.; Dos Santos Pereira Andrade, A.C.; Araújo Lima Rodrigues, R.; La Scola, B.; Santos Abrahão, J. The Multiple Origins of Proteins Present in Tupanvirus Particles. *Curr. Opin. Virol.* **2019**, *36*, 25–31. [CrossRef] [PubMed]
36. Raoult, D. The Post-Darwinist Rhizome of Life. *Lancet* **2010**, *375*, 104–105. [CrossRef]
37. Abrahão, J.S.; Araújo, R.; Colson, P.; Scola, B.L. The Analysis of Translation-Related Gene Set Boosts Debates around Origin and Evolution of Mimiviruses. *PLoS Genet.* **2017**, *13*, e1006532. [CrossRef]
38. Aherfi, S.; La Scola, B.; Pagnier, I.; Raoult, D.; Colson, P. The Expanding Family Marseilleviridae. *Virology* **2014**, *466–467*, 27–37. [CrossRef]
39. Thomas, V.; Bertelli, C.; Collyn, F.; Casson, N.; Telenti, A.; Goesmann, A.; Croxatto, A.; Greub, G. Lausannevirus, a Giant Amoebal Virus Encoding Histone Doublets. *Environ. Microbiol.* **2011**, *13*, 1454–1466. [CrossRef]
40. Aherfi, S.; Pagnier, I.; Fournous, G.; Raoult, D.; La Scola, B.; Colson, P. Complete Genome Sequence of Cannes 8 Virus, a New Member of the Proposed Family "Marseilleviridae". *Virus Genes* **2013**, *47*, 550–555. [CrossRef]
41. Pagnier, I.; Reteno, D.-G.I.; Saadi, H.; Boughalmi, M.; Gaia, M.; Slimani, M.; Ngounga, T.; Bekliz, M.; Colson, P.; Raoult, D.; et al. A Decade of Improvements in Mimiviridae and Marseilleviridae Isolation from Amoeba. *Intervirology* **2013**, *56*, 354–363. [CrossRef]
42. Aherfi, S.; Boughalmi, M.; Pagnier, I.; Fournous, G.; La Scola, B.; Raoult, D.; Colson, P. Complete Genome Sequence of Tunisvirus, a New Member of the Proposed Family Marseilleviridae. *Arch. Virol.* **2014**, *159*, 2349–2358. [CrossRef] [PubMed]
43. Boughalmi, M.; Pagnier, I.; Aherfi, S.; Colson, P.; Raoult, D.; La Scola, B. First Isolation of a Marseillevirus in the Diptera Syrphidae Eristalis Tenax. *Intervirology* **2013**, *56*, 386–394. [CrossRef]
44. Lagier, J.-C.; Armougom, F.; Million, M.; Hugon, P.; Pagnier, I.; Robert, C.; Bittar, F.; Fournous, G.; Gimenez, G.; Maraninchi, M.; et al. Microbial Culturomics: Paradigm Shift in the Human Gut Microbiome Study. *Clin. Microbiol. Infect.* **2012**, *18*, 1185–1193. [CrossRef]
45. Colson, P.; Fancello, L.; Gimenez, G.; Armougom, F.; Desnues, C.; Fournous, G.; Yoosuf, N.; Million, M.; La Scola, B.; Raoult, D. Evidence of the Megavirome in Humans. *J. Clin. Virol.* **2013**, *57*, 191–200. [CrossRef] [PubMed]
46. Doutre, G.; Philippe, N.; Abergel, C.; Claverie, J.-M. Genome Analysis of the First Marseilleviridae Representative from Australia Indicates That Most of Its Genes Contribute to Virus Fitness. *J. Virol.* **2014**, *88*, 14340–14349. [CrossRef]
47. Dornas, F.P.; Assis, F.L.; Aherfi, S.; Arantes, T.; Abrahão, J.S.; Colson, P.; La Scola, B. A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. *Viruses* **2016**, *8*, 76. [CrossRef] [PubMed]
48. Edgar, R.C. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinform.* **2004**, *5*, 113. [CrossRef] [PubMed]
49. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
50. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
51. Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
52. Dos Santos, R.N.; Campos, F.S.; Medeiros de Albuquerque, N.R.; Finoketti, F.; Côrrea, R.A.; Cano-Ortiz, L.; Assis, F.L.; Arantes, T.S.; Roehe, P.M.; Franco, A.C. A New Marseillevirus Isolated in Southern Brazil from Limnoperna Fortunei. *Sci. Rep.* **2016**, *6*, 35237. [CrossRef]
53. Philippe, N.; Legendre, M.; Doutre, G.; Couté, Y.; Poirot, O.; Lescot, M.; Arslan, D.; Seltzer, V.; Bertaux, L.; Bruley, C.; et al. Pandoraviruses: Amoeba Viruses with Genomes up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* **2013**, *341*, 281–286. [CrossRef]
54. Scheid, P.; Balczun, C.; Schaub, G.A. Some Secrets Are Revealed: Parasitic Keratitis Amoebae as Vectors of the Scarcely Described Pandoraviruses to Humans. *Parasitol. Res.* **2014**, *113*, 3759–3764. [CrossRef] [PubMed]

55. Antwerpen, M.H.; Georgi, E.; Zoeller, L.; Woelfel, R.; Stoecker, K.; Scheid, P. Whole-Genome Sequencing of a Pandoravirus Isolated from Keratitis-Inducing Acanthamoeba. *Genome Announc.* **2015**, *3*, e00136-15. [CrossRef]
56. Legendre, M.; Alempic, J.-M.; Philippe, N.; Lartigue, A.; Jeudy, S.; Poirot, O.; Ta, N.T.; Nin, S.; Couté, Y.; Abergel, C.; et al. Pandoravirus Celtis Illustrates the Microevolution Processes at Work in the Giant Pandoraviridae Genomes. *Front. Microbiol.* **2019**, *10*, 430. [CrossRef]
57. Legendre, M.; Fabre, E.; Poirot, O.; Jeudy, S.; Lartigue, A.; Alempic, J.-M.; Beucher, L.; Philippe, N.; Bertaux, L.; Christo-Foroux, E.; et al. Diversity and Evolution of the Emerging Pandoraviridae Family. *Nat. Commun.* **2018**, *9*, 2285. [CrossRef]
58. Pereira Andrade, A.C.D.S.; Victor de Miranda Boratto, P.; Rodrigues, R.A.L.; Bastos, T.M.; Azevedo, B.L.; Dornas, F.P.; Oliveira, D.B.; Drumond, B.P.; Kroon, E.G.; Abrahão, J.S. New Isolates of Pandoraviruses: Contribution to the Study of Replication Cycle Steps. *J. Virol.* **2019**, *93*, e01942-18. [CrossRef] [PubMed]
59. Andreani, J.; Aherfi, S.; Bou Khalil, J.Y.; Di Pinto, F.; Bitam, I.; Raoult, D.; Colson, P.; La Scola, B. Cedratvirus, a Double-Cork Structured Giant Virus, Is a Distant Relative of Pithoviruses. *Viruses* **2016**, *8*, 300. [CrossRef]
60. Bertelli, C.; Mueller, L.; Thomas, V.; Pillonel, T.; Jacquier, N.; Greub, G. Cedratvirus Lausannensis—Digging into Pithoviridae Diversity. *Environ. Microbiol.* **2017**, *19*, 4022–4034. [CrossRef]
61. Jeudy, S.; Rigou, S.; Alempic, J.-M.; Claverie, J.-M.; Abergel, C.; Legendre, M. The DNA Methylation Landscape of Giant Viruses. *Nat. Commun.* **2020**, *11*, 2657. [CrossRef]
62. Andreani, J.; Khalil, J.Y.B.; Baptiste, E.; Hasni, I.; Michelle, C.; Raoult, D.; Levasseur, A.; La Scola, B. Orpheovirus IHUMI-LCC2: A New Virus among the Giant Viruses. *Front. Microbiol.* **2018**, *8*, 2643. [CrossRef]
63. Rodrigues, R.A.L.; Andreani, J.; Andrade, A.C.D.S.P.; Machado, T.B.; Abdi, S.; Levasseur, A.; Abrahão, J.S.; La Scola, B. Morphologic and Genomic Analyses of New Isolates Reveal a Second Lineage of Cedratviruses. *J. Virol.* **2018**, *92*, e00372-18. [CrossRef]
64. Souza, F.; Rodrigues, R.; Reis, E.; Lima, M.; La Scola, B.; Abrahão, J. In-Depth Analysis of the Replication Cycle of Orpheovirus. *Virol. J.* **2019**, *16*, 158. [CrossRef]
65. Reteno, D.G.; Benamar, S.; Khalil, J.B.; Andreani, J.; Armstrong, N.; Klose, T.; Rossmann, M.; Colson, P.; Raoult, D.; La Scola, B. Faustovirus, an Asfarvirus-Related New Lineage of Giant Viruses Infecting Amoebae. *J. Virol.* **2015**, *89*, 6585–6594. [CrossRef] [PubMed]
66. Oliveira, G.P.; de Aquino, I.L.M.; Luiz, A.P.M.F.; Abrahão, J.S. Putative Promoter Motif Analyses Reinforce the Evolutionary Relationships Among Faustoviruses, Kaumoebavirus, and Asfarvirus. *Front. Microbiol.* **2018**, *9*, 1041. [CrossRef] [PubMed]
67. Tully, B.J.; Graham, E.D.; Heidelberg, J.F. The Reconstruction of 2,631 Draft Metagenome-Assembled Genomes from the Global Oceans. *Sci. Data* **2018**, *5*, 170203. [CrossRef]
68. Schulz, F.; Roux, S.; Paez-Espino, D.; Jungbluth, S.; Walsh, D.A.; Denef, V.J.; McMahon, K.D.; Konstantinidis, K.T.; Eloe-Fadrosh, E.A.; Kyrpides, N.C.; et al. Giant Virus Diversity and Host Interactions through Global Metagenomics. *Nature* **2020**, *578*, 432–436. [CrossRef]
69. Hingamp, P.; Grimsley, N.; Acinas, S.G.; Clerissi, C.; Subirana, L.; Poulain, J.; Ferrera, I.; Sarmento, H.; Villar, E.; Lima-Mendez, G.; et al. Exploring Nucleo-Cytoplasmic Large DNA Viruses in Tara Oceans Microbial Metagenomes. *ISME J.* **2013**, *7*, 1678–1695. [CrossRef] [PubMed]
70. Rodrigues, R.A.L.; dos Santos Silva, L.K.; Dornas, F.P.; de Oliveira, D.B.; Magalhães, T.F.F.; Santos, D.A.; Costa, A.O.; de Macêdo Farias, L.; Magalhães, P.P.; Bonjardim, C.A.; et al. Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. *J. Virol.* **2015**, *89*, 11812–11819. [CrossRef]
71. Ghigo, E.; Kartenbeck, J.; Lien, P.; Pelkmans, L.; Capo, C.; Mege, J.-L.; Raoult, D. Ameobal Pathogen Mimivirus Infects Macrophages through Phagocytosis. *PLoS Pathog.* **2008**, *4*, e1000087. [CrossRef]
72. Andrade, A.C.D.S.P.; Rodrigues, R.A.L.; Oliveira, G.P.; Andrade, K.R.; Bonjardim, C.A.; La Scola, B.; Kroon, E.G.; Abrahão, J.S. Filling Knowledge Gaps for Mimivirus Entry, Uncoating, and Morphogenesis. *J. Virol.* **2017**, *91*, e01335-17. [CrossRef] [PubMed]
73. Schrad, J.R.; Abrahão, J.S.; Cortines, J.R.; Parent, K.N. Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. *Cell* **2020**, *181*, 1046–1061.e6. [CrossRef] [PubMed]
74. De Souza, G.A.P.; Queiroz, V.F.; Coelho, L.F.L.; Abrahão, J.S. Alohomora! What the Entry Mechanisms Tell Us about the Evolution and Diversification of Giant Viruses and Their Hosts. *Curr. Opin. Virol.* **2021**, *47*, 79–85. [CrossRef] [PubMed]
75. Quemin, E.R.; Corroyer-Dulmont, S.; Krijnse-Locker, J. Entry and Disassembly of Large DNA Viruses: Electron Microscopy Leads the Way. *J. Mol. Biol.* **2018**, *430*, 1714–1724. [CrossRef]
76. Parent, K.N.; Schrad, J.R.; Young, E.J.; Abrahao, J.S.; Cortines, J.R. A Gateway into Understanding the Unique Vertex of Samba Virus. *Microsc. Microanal.* **2018**, *24*, 1438–1439. [CrossRef]
77. Zauberman, N.; Mutsafi, Y.; Halevy, D.B.; Shimoni, E.; Klein, E.; Xiao, C.; Sun, S.; Minsky, A. Distinct DNA Exit and Packaging Portals in the Virus Acanthamoeba Polyphaga Mimivirus. *PLoS Biol.* **2008**, *6*, e114. [CrossRef] [PubMed]
78. Oliveira, G.; Silva, L.; Leão, T.; Mougari, S.; da Fonseca, F.G.; Kroon, E.G.; La Scola, B.; Abrahão, J.S. Tupanvirus-Infected Amoebas Are Induced to Aggregate with Uninfected Cells Promoting Viral Dissemination. *Sci. Rep.* **2019**, *9*, 183. [CrossRef] [PubMed]
79. De Farias, S.T.; Jheeta, S.; Prosdocimi, F. Viruses as a Survival Strategy in the Armory of Life. *Hist. Philos. Life Sci.* **2019**, *41*, 45. [CrossRef] [PubMed]
80. Oliveira, G.; La Scola, B.; Abrahão, J. Giant Virus vs Amoeba: Fight for Supremacy. *Virol. J.* **2019**, *16*, 126. [CrossRef]

81. Boratto, P.; Albarnaz, J.D.; Almeida, G.M.; Botelho, L.; Fontes, A.C.L.; Costa, A.O.; Santos, D.d.A.; Bonjardim, C.A.; La Scola, B.; Kroon, E.G.; et al. Acanthamoeba Polyphaga Mimivirus Prevents Amoebal Encystment-Mediating Serine Proteinase Expression and Circumvents Cell Encystment. *J. Virol.* **2015**, *89*, 2962–2965. [CrossRef]
82. Silva, L.K.D.S.; Boratto, P.V.M.; La Scola, B.; Bonjardim, C.A.; Abrahão, J.S. Acanthamoeba and Mimivirus Interactions: The Role of Amoebal Encystment and the Expansion of the "Cheshire Cat" Theory. *Curr. Opin. Microbiol.* **2016**, *31*, 9–15. [CrossRef] [PubMed]
83. Korn, E.D.; Weisman, R.A. Phagocytosis of Latex Beads by Acanthamoeba. II. Electron Microscopic Study of the Initial Events. *J. Cell Biol.* **1967**, *34*, 219–227. [CrossRef] [PubMed]
84. Arantes, T.S.; Rodrigues, R.A.L.; Dos Santos Silva, L.K.; Oliveira, G.P.; de Souza, H.L.; Khalil, J.Y.B.; de Oliveira, D.B.; Torres, A.A.; da Silva, L.L.; Colson, P.; et al. The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles. *J. Virol.* **2016**, *90*, 5246–5255. [CrossRef]
85. Oliveira, G.P.; Lima, M.T.; Arantes, T.S.; Assis, F.L.; Rodrigues, R.A.L.; da Fonseca, F.G.; Bonjardim, C.A.; Kroon, E.G.; Colson, P.; La Scola, B.; et al. The Investigation of Promoter Sequences of Marseilleviruses Highlights a Remarkable Abundance of the AAATATTT Motif in Intergenic Regions. *J. Virol.* **2017**, *91*, e01088-17. [CrossRef] [PubMed]
86. Rodrigues, R.A.L.; Louazani, A.C.; Picorelli, A.; Oliveira, G.P.; Lobo, F.P.; Colson, P.; La Scola, B.; Abrahão, J.S. Analysis of a Marseillevirus Transcriptome Reveals Temporal Gene Expression Profile and Host Transcriptional Shift. *Front. Microbiol.* **2020**, *11*, 651. [CrossRef]
87. Serrano-Solís, V.; Toscano Soares, P.E.; de Farías, S.T. Genomic Signatures Among Acanthamoeba Polyphaga Entoorganisms Unveil Evidence of Coevolution. *J. Mol. Evol.* **2019**, *87*, 7–15. [CrossRef]
88. Akashi, M.; Fukaya, S.; Uchiyama, C.; Aoki, K.; Takemura, M. Visualization of Giant Virus Particles and Development of "VIRAMOS" for High School and University Biology Course. *Biochem. Mol. Biol. Educ.* **2019**, *47*, 426–431. [CrossRef]
89. Paez-Espino, D.; Eloe-Fadrosh, E.A.; Pavlopoulos, G.A.; Thomas, A.D.; Huntemann, M.; Mikhailova, N.; Rubin, E.; Ivanova, N.N.; Kyrpides, N.C. Uncovering Earth's Virome. *Nature* **2016**, *536*, 425–430. [CrossRef]
90. Matza-Porges, S.; Nathan, D. A Biosafety Level 2 Virology Lab for Biotechnology Undergraduates. *Biochem. Mol. Biol. Educ.* **2017**, *45*, 537–543. [CrossRef]
91. De Souza, G.A.P.; Queiroz, V.F.; Lima, M.T.; de Sousa Reis, E.V.; Coelho, L.F.L.; Abrahão, J.S. Virus Goes Viral: An Educational Kit for Virology Classes. *Virol. J.* **2020**, *17*, 13. [CrossRef] [PubMed]

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Diversity of Surface Fibril Patterns in Mimivirus Isolates

Isabella Luiza Martins de Aquino,[a] Mateus Sá Magalhães Serafim,[a] Talita Bastos Machado,[a] Bruna Luiza Azevedo,[a] Denilson Eduardo Silva Cunha,[b] Leila Sabrina Ullmann,[c] João Pessoa Araújo, Jr.,[c] Jônatas Santos Abrahão[a]

[a]Laboratório de Vírus, Instituto de Ciências Biológicas, Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[b]Centro de Microscopia da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[c]Laboratório de Virologia, Departamento de Microbiologia e Imunologia, Instituto de Biotecnologia, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil

ABSTRACT Among the most intriguing structural features in the known virosphere are mimivirus surface fibrils, proteinaceous filaments approximately 150 nm long, covering the mimivirus capsid surface. Fibrils are important to promote particle adhesion to host cells, triggering phagocytosis and cell infection. However, although mimiviruses are one of the most abundant viral entities in a plethora of biomes worldwide, there has been no comparative analysis on fibril organization and abundance among distinct mimivirus isolates. Here, we describe the isolation and characterization of Megavirus caiporensis, a novel lineage C mimivirus with surface fibrils organized as "clumps." This intriguing feature led us to expand our analyses to other mimivirus isolates. By employing a combined approach including electron microscopy, image processing, genomic sequencing, and viral prospection, we obtained evidence of at least three main patterns of surface fibrils that can be found in mimiviruses: (i) isolates containing particles with abundant fibrils, distributed homogeneously on the capsid surface; (ii) isolates with particles almost fibrilless; and (iii) isolates with particles containing fibrils in abundance, but organized as clumps, as observed in Megavirus caiporensis. A total of 15 mimivirus isolates were analyzed by microscopy, and their DNA polymerase subunit B genes were sequenced for phylogenetic analysis. We observed a unique match between evolutionarily-related viruses and their fibril profiles. Biological assays suggested that patterns of fibrils can influence viral entry in host cells. Our data contribute to the knowledge of mimivirus fibril organization and abundance, as well as raising questions on the evolution of those intriguing structures.

IMPORTANCE Mimivirus fibrils are intriguing structures that have drawn attention since their discovery. Although still under investigation, the function of fibrils may be related to host cell adhesion. In this work, we isolated and characterized a new mimivirus, called Megavirus caiporensis, and we showed that mimivirus isolates can exhibit at least three different patterns related to fibril organization and abundance. In our study, evolutionarily-related viruses presented similar fibril profiles, and such fibrils may affect how those viruses trigger phagocytosis in amoebas. These data shed light on aspects of mimivirus particle morphology, virus-host interactions, and their evolution.

KEYWORDS amoeba, cell adhesion, diversity, fibrils, mimivirus, structural biology, virus entry

In 2003, *Acanthamoeba polyphaga* mimivirus (APMV) was the first amoeba-associated giant virus to be described and studied (1). The APMV exhibits particles composed of a capsid with pseudo-icosahedral symmetry covered by a dense layer of external fibrils, approximately 750 nm in size. The capsid is formed by several protein layers and an inner lipid membrane, which encompasses the viral core, where the viral genome is located (2). Recently, a study suggested that the APMV genome is organized within a 30-nm helical protein shell composed primarily of 2-glucose-methanol-choline (GMC)

Editor Derek Walsh, Northwestern University Feinberg School of Medicine

Address correspondence to Jônatas Santos Abrahão, jonatas.abrahao@gmail.com.

The authors declare no conflict of interest.

Received 23 November 2022
Accepted 7 December 2022

oxidoreductases (genomic fiber) encoded by the R135 gene, which was also shown to be one of the most abundant elements of the surface fibrils that cover the capsid (3, 4). Fibrils are important structures for adhesion to the surface of amoebas (5). For instance, APMV fibrils are often found with one of their ends free and associated with a globular terminal, while the other is attached to a single central structure (6). In addition, they appear to be associated with a peptidoglycan-like structure, exhibiting successive rings of density (6). Moreover, in addition to R135, the L829 and L725 proteins have also been associated with APMV fibril composition, as the lack of those genes in the genome of mimivirus M4 was linked to an almost-fibrilless phenotype of M4 isolate particles (7). To date, the M4 isolate was obtained after successive passages of APMV in amoebas under allopatric conditions, resulting in genome shrinking and a decrease in fibril abundance (7). Furthermore, the adhesion of mimivirus fibrils to the host membrane seemed to be mediated by sugars found on both the host surface and in the fibrils (e.g., mannose and *N*-acetylglucosamine), allowing the phagocytosis of the virus and subsequent entry to start the cycle (5). Additionally, mimiviruses have a star-shaped structure called a stargate at one of the vertices of their particle, and this structure is responsible for releasing the genome at the beginning of the cycle within the host and is the only region at the capsid uncovered by fibrils (2, 6).

Currently, mimiviruses belong to the genus *Mimivirus*, family *Mimiviridae*, order *Imitervirales*, class *Megaviricetes*, and phylum *Nucleocytoviricota* (8). Metagenomics studies indicate that mimiviruses are among the most abundant viral entities in different biomes worldwide (9, 10). Phylogenetic analyses considering hallmark genes have shown that mimiviruses seem to be divided into three main lineages, A, B, and C (11). Lineage A is represented by APMV and other viruses similar to it, such as the Brazilian isolate sambavirus (12). Lineage B consists of the moumouviruses and related viruses, and lineage C is composed of megaviruses such as Megavirus chilensis and other similar viruses (13, 14). Genetic analyses show that mimivirus lineages, despite being evolutionarily related, present important differences between some of their genes. In this context, the possible differences in the morphological organization of mimivirus fibrils have not been explored in a comparative way, considering the three lineages and/or their isolates, and is an open field for study and new discoveries.

In this work, we describe the isolation of a novel mimivirus belonging to lineage C, named Megavirus caiporensis. The microscopic and genomic analyses of this virus raised questions on how surface fibrils are organized in different mimivirus isolates and encouraged us to expand this study to 14 other mimivirus isolates. Taken together, our data demonstrate at least three main patterns of fibril organization among isolates from different lineages. Considering the viruses analyzed here, evolutionarily-related isolates present similar fibril profiles. Our data also indicate that, in contrast to previous speculations (7), even the almost-fibrilless isolates (e.g., Borely moumouvirus) can have the R135, L829, and L725 genes in their genomes.

RESULTS

Megavirus caiporensis: isolation of a new mimivirus in Brazil. As part of our continuous efforts to discover new giant viruses of amoebas, a new virus was isolated from a water sample collected from an urban lagoon in Belo Horizonte, Brazil. After observing cytopathic effect in the microplate well where this sample was inoculated, we proceeded with characterization of this isolate. By light microscopy, we observed the development and progression of the infection cycle as cytopathic effects emerged and intensified. At 6 h postinfection (hpi), the cells began to round and detach from the monolayer. At 12 hpi, fully rounded cells began to undergo lysis, and at 24 hpi almost all cells were lysed and cellular debris was observed in the supernatant and at the bottom of the flasks (Fig. 1A).

Images obtained by transmission and scanning electron microscopy (TEM and SEM, respectively) allowed us to identify the isolate as a mimivirus (Fig. 1B and C), due to the morphologic characteristics of its particles, composed of capsids with an average

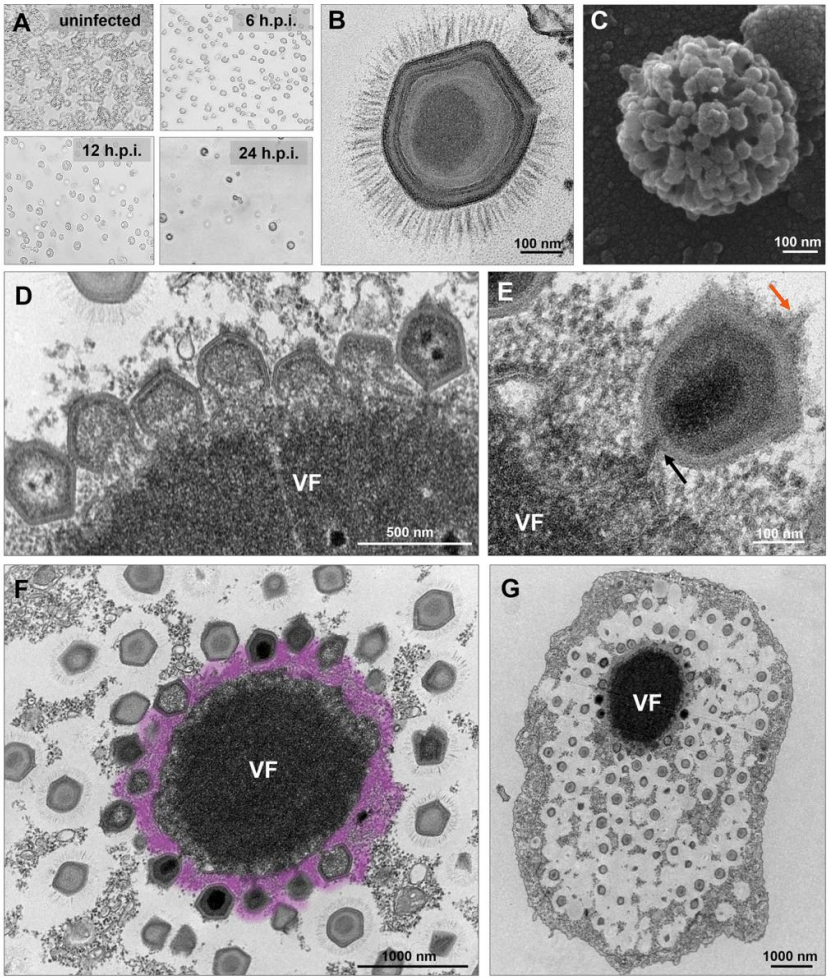

FIG 1 Megavirus caiporensis viral particle and cycle. (A) Cytopathic effects on *A. castellanii* amoeba cells infected with Megavirus caiporensis. For the uninfected cell control, it is possible to observe the cells presenting irregular shapes, vacuoles, and good adhesion forming a monolayer, characteristics that allowed us to verify a healthy culture. From 6 h postinfection (hpi) with the megavirus that we isolated, the cells had already begun to undergo rounding and loosening of the monolayer, as seen in abundance in the supernatant. After 12 hpi, complete rounding and the beginning of cell lysis were observed, which significantly increased after 24 hpi, at which time few cells were still intact. Images were obtained using 100× magnification. (B and C) Characteristics of the Megavirus caiporensis viral particle. The Megavirus caiporensis capsid is composed of multiple layers involving the genome, a central and darker region, and a layer of fibrils covering the structure (B). The SEM image of a Megavirus caiporensis particle is shown (C). (D to G) Megavirus caiporensis multiplication cycle stages. (D) Assembly of new Megavirus caiporensis particles, beginning with the filling of the crescent-shaped form with the material present in the viral factory (VF). (E) Megavirus caiporensis viral particle receiving the internal contents through the opposite end (black arrow) to the stargate (orange arrow). (F) Viral factory producing new particles and the fibril acquisition area highlighted in purple. (G) The mature viral factory (VF) occupies a large part of the host cytoplasm of infected amoebas in full production of new Megavirus caiporensis particles.

size of 435 nm, covered by a layer of fibrils of 108 nm, and a total size of 651 nm (average). Interestingly, compared to what has been described for APMV (6), this isolate exhibited fibrils in a different organization, forming small clumps (Fig. 1B and C). Fibrils seemed to be grouped from the point of insertion into the capsid to their outermost portions (Fig. 1B). Here, as the fibrils of this virus are similar to those observed for

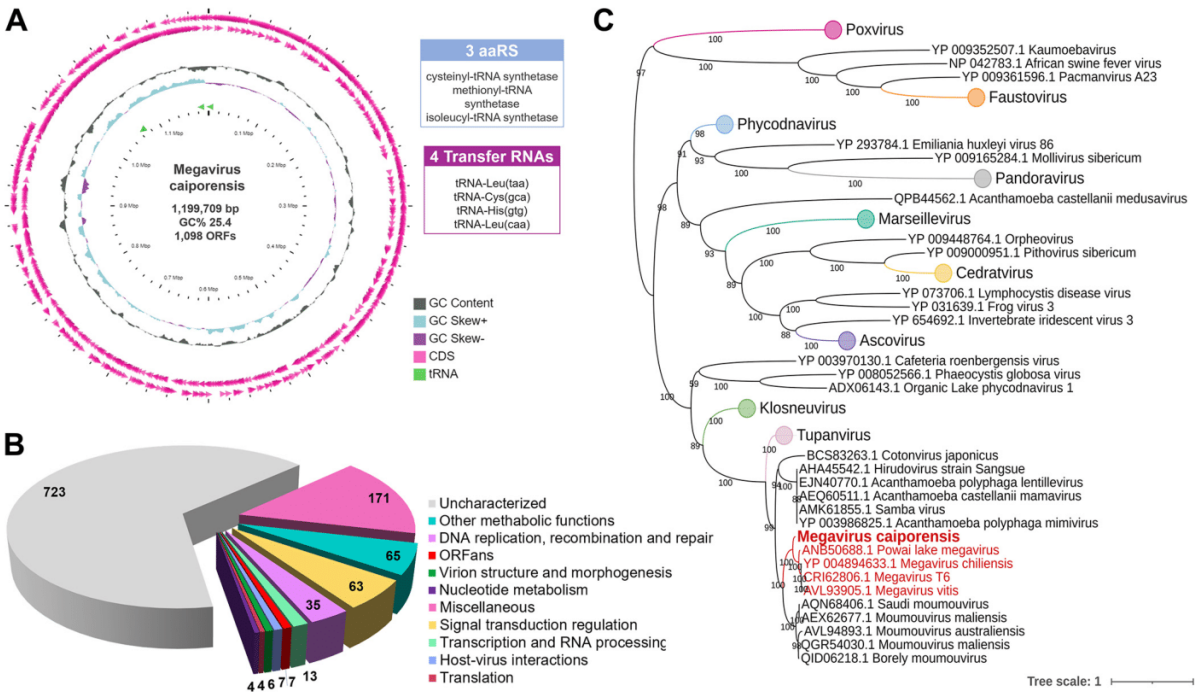

FIG 2 Genomic features of Megavirus caiporensis. (A) Circular representation of the Megavirus caiporensis genome. Rings, from innermost to outermost, correspond to genome coordinates in kilobases (kb), GC skew, GC content, and predicted protein-coding sequences (CDS) on both the forward and reverse strands. The green arrowheads correspond to transfer RNAs. A colored subtitle is provided on the right side of the figure. The blue and purple boxes at right indicate the three aminoacyl-tRNA synthetases (aaRS) and four tRNAs identified for this virus, respectively. (B) A set of genes from Megavirus caiporensis, classified according predicted gene categories. A colored subtitle is provided on the right side of the figure. (C) Phylogenetic analysis of subfamily B DNA polymerase sequences of Megavirus caiporensis and *Nucleocytoviricota*-related viruses. A maximum-likelihood phylogenetic tree was constructed with sequences of subfamily B DNA polymerase of mimiviruses and other *Nucleocytoviricota*-associated viruses. In this phylogeny, the new isolate described here, Megavirus caiporensis, clustered with lineage C mimiviruses (highlighted in red). The scale bar indicates the rate of evolution.

Megavirus chilensis (13) and considering the sequencing results on this isolate (discussed below), we named it Megavirus caiporensis.

TEM images revealed that the entire cycle of Megavirus caiporensis is similar to what has been previously described for other mimiviruses (15). For instance, the immature viral factory is formed in the amoeba cytoplasm after virus entry, likely by phagocytosis, initiating the replication cycle. After the formation and maturation of the viral factory, mitochondria can be observed in its periphery, as well as formation of new particles assembling by crescent-shape structures, which gradually increase in size and are filled by the content present in the factory (Fig. 1D and E). During the acquisition of fibrils at the so-called fibril acquisition area, the particles receive their internal content from the opposite side of the stargate (Fig. 1D to F). At the end of the cycle, mature factories occupy a large area in the cell cytoplasm, as described for other mimiviruses (15). The newly formed particles agglomerate into the host cytoplasm, and the factories decrease in size throughout the replication cycle (Fig. 1G). Finally, particle release occurs by cell lysis.

Megavirus caiporensis sequencing and phylogenetic analysis confirmed it belongs to mimivirus lineage C. The genome of Megavirus caiporensis is a linear, double-stranded DNA molecule of 1,199,709 bp in length (Fig. 2A). Its genomic content is predicted to encode 1,098 genes (568 located on the negative strand and 530 located on the positive strand) with a coding density of 91.3% and a GC content of 25.4%. Among the predicted proteins, 723 had no known function and were considered uncharacterized. Most functional genes fit into the categories miscellaneous (i.e., domain and repeat proteins; diverse enzymes) (171), other metabolic functions (65), signal transduction regulation (63), and DNA recombination, replication, and repair (35) (Fig. 2B). B-family DNA polymerase, chaperone, helicases, and replication factors

are some of the genes that are incorporated inside the category of DNA recombination, replication, and repair. Seven open reading frames (ORFs) with no detectable homology to other ORFs in a database were detected and were considered ORFans. In addition, genes related to translation, a trademark of mimiviruses, were identified, including three aminoacyl-tRNA-synthetases, four RNA transporters (Fig. 2A), and one GTP binding translation elongation/initiation factor. The Megavirus caiporensis genome presents 13 genes related to transcription and RNA processing, including RNA ligase and different subunits of DNA-directed RNA polymerase. The majority of genes predicted matched those also predicted for Powai Lake megavirus and Megavirus chilensis, including ankyrin repeat proteins (miscellaneous category), putative lipoprotein (other metabolic functions), and putative protein kinases (signal transduction regulation category). Genes associated with host-virus interactions, nucleotide metabolism, and virion structure and morphogenesis were also predicted for Megavirus caiporensis. Additionally, in order to investigate the evolutionary and phylogenetic relationship of Megavirus caiporensis with other *Nucleocytoviricota*, including mimiviruses, we also performed a phylogenetic analysis using the conserved gene that encodes family B DNA polymerase as a target (Fig. 2C). Phylogenetic construction confirmed that Megavirus caiporensis clustered with lineage C mimiviruses.

Mimivirus isolates have at least three patterns of surface fibrils. The isolation and characterization of Megavirus caiporensis intrigued us regarding its structural aspects, as we noticed a different organization of its fibrils, further expanding these analyses to lineage A and B mimiviruses. Thus, we selected APMV (lineage A) and Borely moumouvirus (lineage B), as they are two viruses for which the whole genomes have been sequenced, as well as their being available in our laboratory. In order to analyze those isolates' fibrils, particles of APMV, Borely moumouvirus, and Megavirus caiporensis were observed by TEM and SEM (Fig. 3A). As for APMV, it was possible to observe, as previously described in the literature, that its fibrils are long and abundantly surround the capsid (Fig. 3A), herein termed pattern A. In comparison, Borely moumouvirus (lineage B) appeared to have fewer fibrils around its capsid and these were less homogeneously distributed (Fig. 3A); here, we termed this pattern B. Interestingly, some particles belonging to pattern B were almost fibrilless. Finally, the isolate described here, Megavirus caiporensis, which belongs to lineage C, had an abundant number of fibrils, as seen for lineage A, but they were organized in small groups, as clumps (Fig. 3A); this pattern is herein termed pattern C. In SEM images, the patterns described here for fibrils were maintained, and we also observed very clearly the aforementioned differences among these mimiviruses (Fig. 3B). For instance, APMV exhibited its surface with fibrils homogeneously distributed, while for Megavirus caiporensis fibrils appeared to be organized in groups or clusters. As for Borely moumouvirus, it is possible to see that the capsid surface exhibits a geometric appearance, which probably is linked with the lower number of fibrils that is usually observed in EM images. We also developed a protocol to estimate the relative abundance of surface fibrils in the particles of mimiviruses, and then we applied it to compare APMV, Borely moumouvirus, and Megavirus caiporensis (Fig. 3C). After calibration of average values, the analysis revealed that Megavirus caiporensis exhibited the highest relative average contrast, 552-fold higher than Borely moumouvirus, followed by APMV, for which the contrast was 394-fold higher than Borely moumouvirus (Fig. 3D). Megavirus caiporensis had the highest relative average contrast and this was possibly related not only to the number of fibrils but also to the clump's organization, while Borely moumouvirus having the lowest contrast was expected, since it has fewer fibrils surrounding the capsid (Fig. 3D).

It is important to mention that preparation for TEM can influence fibril appearance. To understand whether the different fibril organizations and abundances could be the result of possible artifacts of preparation, we concomitantly analyzed different combinations of mixed purified particles of APMV, Borely moumouvirus, and Megavirus caiporensis by TEM. Therefore, those different viruses were mixed in the same sample in

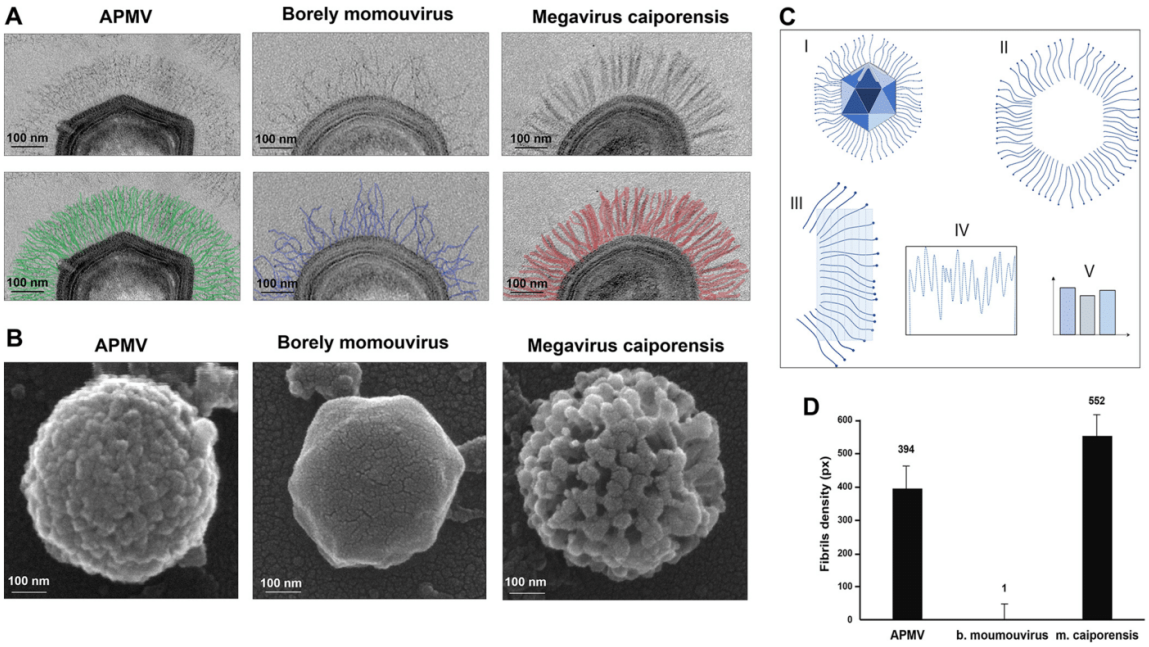

FIG 3 Morphological differences of mimivirus fibrils according to lineages A, B, and C. (A) Different patterns of fibril organization observed by TEM. Mimivirus extern fibrils in TEM images are highlighted with different colors to show the discrepancies of the fibril abundances and organizations (APMV in green, Borely moumouvirus in blue, and Megavirus caiporensis in red). (B) Different patterns of fibril organization observed by SEM. SEM images show the fibrils on the surfaces of the particles of APMV, Borely moumouvirus, and Megavirus caiporensis, respectively. Scale bars are indicated for each image. (C) Analysis of mimivirus fibril densities. For this illustrative representation of the protocol used to analyze the densities of fibrils of different mimiviruses, we have provided images obtained by TEM for each mimivirus analyzed (I). The fibrils were highlighted in the images manually by using the drawing tool available in Microsoft PowerPoint version 2021, and the background image was erased, leaving only the fibrils that were highlighted. (III) For each of the six sides of the particles, the fibril area was covered with rectangles of the same size and thickness in ImageJ software (version v1.53k). (IV) From each rectangle we generated a graph showing curves referring to pixel values read within the demarcated region. (V) The values obtained were plotted in the form of a table in Excel version 2021 and analyzed. (D) Graph representing the densities of fibrils found for each of the mimiviruses analyzed in this work. The *x* axis represents the analyzed lineages and the *y* axis shows the densities of fibrils in pixels. Values were calibrated based on the lowest observed value. Megavirus caiporensis presented the highest value identified among the three (552 pixels), followed by APMV (394 pixels). Finally, Borely moumouvirus presented the lowest number among the three (1 pixel), probably due to its lower number of these structures.

trio or pairs (Fig. 4) and then analyzed. All three fibril patterns were observed in the trio preparation (lineages A + B + C) (Fig. 4A). Patterns A and B were observed in preparations containing a mix of particles belonging to lineages A and B (Fig. 4B), and no particles with clumped fibrils were observed. Patterns B and C were the only ones observed in preparations containing lineage B and C viruses (Fig. 4C). Finally, patterns A and C were the only ones observed in preparations containing lineage A and C viruses (Fig. 4D). Taken together, these results confirmed that the distinct patterns of fibril organization and abundances were not likely to be an artifact of analysis related to TEM sample preparation.

Considering these previous results regarding different mimivirus organization and abundance of fibril patterns, we expanded our analyses to other mimivirus isolates obtained by our group in the past 10 years. To date, all viruses isolated in our lab have been inspected by TEM, and their images are stored in a database. We searched for isolates with the three aforementioned fibril patterns (Fig. 2). We selected 15 isolates, 5 of each pattern (Fig. 5), and sequenced their family B DNA polymerase gene, which is considered a hallmark gene for differentiating the three mimivirus lineages. Sequences were used to construct a data set and a phylogenetic tree considering APMV, Borely moumouvirus, and Megavirus chilensis (13) as references for lineages A, B, and C, respectively. The results revealed a corresponding match among the three lineages and the predicted fibril morphotypes (Fig. 5 and 6): (i) lineage A, fully covered by fibrils (pattern A), including APMV, Mimivirus PU, Mimivirus capivarensis, Kroon mimivirus,

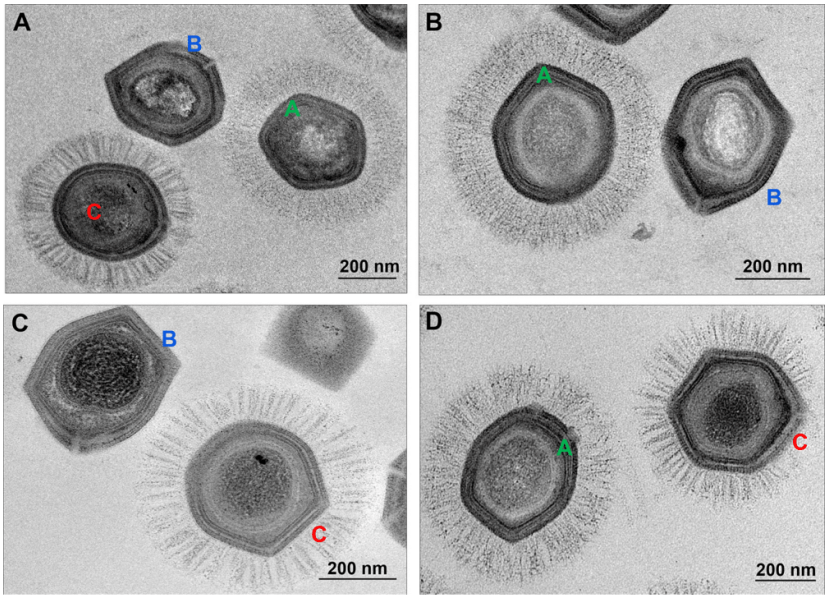

FIG 4 TEM images of trios and pairs of purified mimivirus particles related to lineages A, B, and C obtained in one same sample. (A) Mimivirus trio in the same sample; TEM sample images contain mimivirus A + B + C. (B to D) Mimivirus pairs in the same sample, with TEM sample images for pairs A + B (B), B + C (C), and A + C (D). For the preparations, only purified viral particles were used in combinations in trios or pairs, with APMV representing pattern A; Borely moumouvirus for pattern B; and Megavirus caiporensis representing pattern C. Scale bars are indicated in the images.

and Amazonia mimivirus; (ii) lineage B, almost fibril-less (pattern B), including Borely moumouvirus, Moumouvirus crenensis, Moumouvirus dionensis, Moumouvirus limneidensis, and Moumouvirus naiadiensis; and (iii) lineage C, with fibrils in clumps (pattern C), including Megavirus caiporensis, Megavirus curupirensis, Megavirus botiensis, Megavirus boitataensis, and Megavirus muiraquitaensis. Taken together, our results suggested that, considering our collection of isolates, each lineage is related to a pattern of fibril morphotype. A comprehensive search for TEM images of mimivirus isolates in the literature also revealed that at least two of the three patterns of fibrils presented here have already been observed in natural isolates: those with particles fully covered by fibrils (pattern A) (16–18) and those with clumped fibrils (pattern C) (13, 19–21). Although TEM images of lineage B mimivirus particles are scarce in the literature, there is evidence that some moumouviruses have long abundant fibrils, similar to that described for APMV and other lineage A mimiviruses (22), such as those presented here. Therefore, although our isolates indicated a possible correspondence between patterns of fibrils and mimivirus lineages, this feature may not be expandable to all isolates and requires further investigation.

As aforementioned, the M4 isolate was obtained after successive passages of APMV in amoebas under allopatric conditions, resulting in a different set of phenotypes, including almost fibril-less particles. Proteomic analysis of M4 particles indicated the absence of a GMC-oxidoreductase encoded by the APMV R135 gene, while this gene has been reported as a main component of mimivirus fibrils (4, 6). To evaluate this correlation between fibrils and R135, we investigated the presence of R135 orthologues in the Borely moumouvirus genome. Interestingly, although Borely moumouvirus particles have a pattern similar to the M4 isolate, we found an R135 orthologue in its genome with 68.2% similarity to APMV R135, considering predicted amino acid sequences. Other APMV genes previously reported to be related to fibrils but absent in the M4 genome, such as L829 and L725, were also found in the Borely moumouvirus genome, with 62.9%

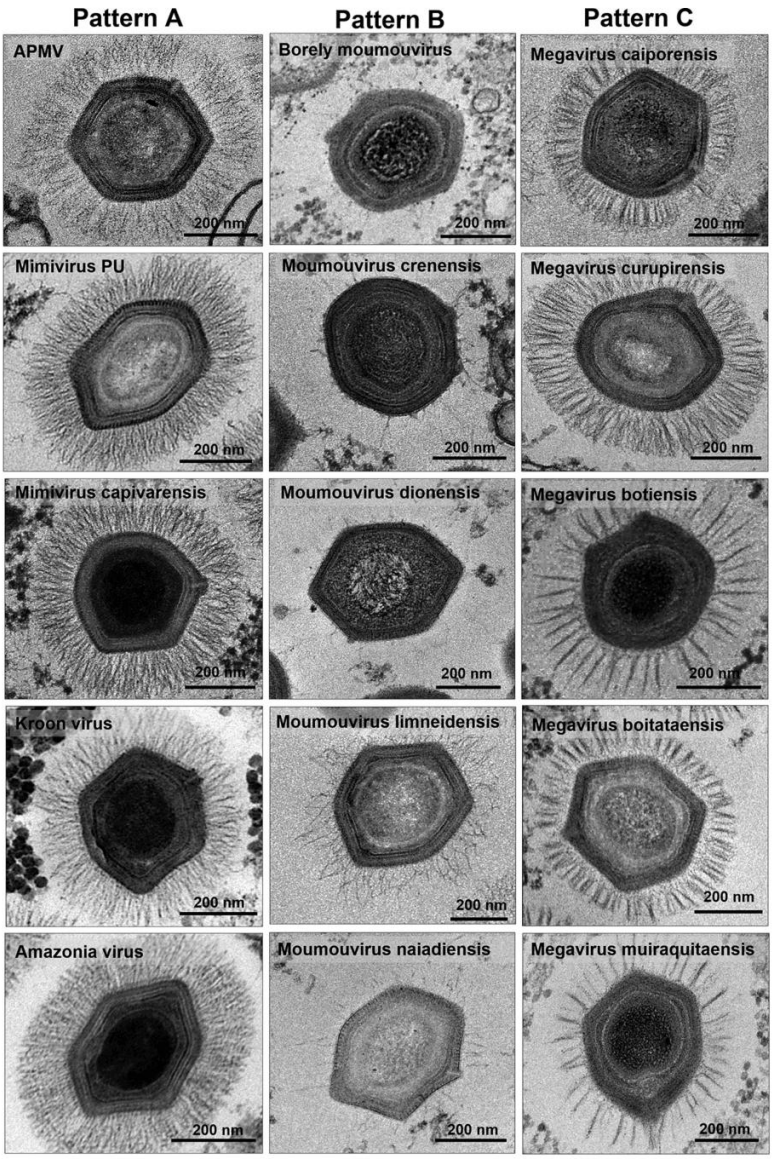

FIG 5 Comparative panel of several mimivirus fibril morphologies related to patterns A, B, and C. Each column represents a pattern of fibril organization for each mimivirus lineage (A, B, or C), exhibited in five different isolates each. Pattern A mimiviruses (first column) fibrils are long and abundantly and homogeneously distributed around the capsid. Pattern B mimiviruses (second column) exhibit many fewer fibrils that are not uniformly distributed, and for some isolates the number of fibrils was too low to be captured by TEM images. Finally, for pattern C mimivirus (third column), the fibrils are organized surrounding the capsid, forming clumps. Some of those fibril clusters are so closely knit that the particle seems to have fewer fibrils (Megavirus muiraquitensis and Megavirus botiensis) compared to others, implicating an interspecific variance between pattern C viruses.

and 54.7% similarity to the APMV predicted proteins, respectively (Table 1). We also observed that those three genes were widespread in most mimiviruses, which were clustered according to their respective lineages (A, B, or C) regardless of their fibrils pattern (Fig. 7A to C).

Fibrils may affect particle incorporation by amoebas. Our group previously demonstrated that mimivirus fibrils may play an important role in viral adhesion to host

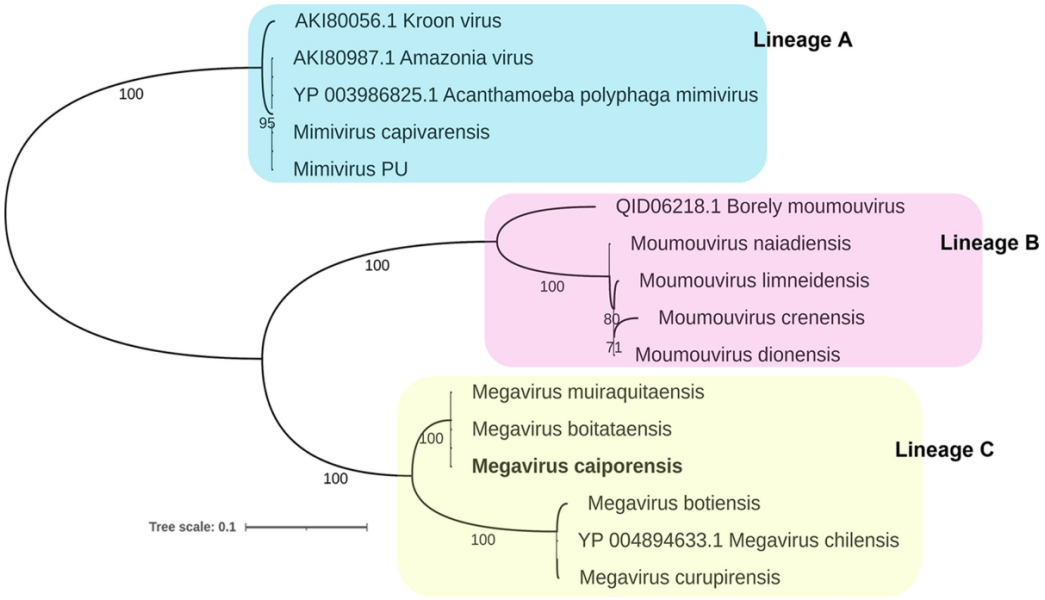

FIG 6 Phylogenetic construction of mimiviruses of patterns A, B, and C. Phylogenetic trees were constructed based on the DNA polymerase gene of subfamily B of 15 mimiviruses, presented previously in a comparative panel of fibril organization patterns. To investigate the history of the relationship of each mimivirus with one lineages, A, B, or C, a representative previously related to each lineage was chosen: APMV for lineage A, Borely moumouvirus for lineage B, and Megavirus chilensis for lineage C. Our results suggested that different fibril patterns are lineage-associated, considering our analyzed isolates.

amoebas (5). To understand whether the pattern of fibrils could affect the adhesion and entry of the particles in *A. castellanii*, we performed a set of biological assays with viruses presenting patterns A, B, and C. Analysis of TEM images showed pattern A and C particles binding to host plasma membrane, mediated by surface fibrils (Fig. 8A), but no pattern B particles were visualized attached to host cells. To quantitatively evaluate particles adhesion to amoebas, cells were infected at a multiplicity of infection (MOI) of 10, and after 60 min the supernatants were collected and titers were determined. The titration of the samples collected 1 hpi revealed approximately 10 times more Borely moumouvirus infectious particles than APMV and Megavirus caiporensis, suggesting that Borely moumouvirus particles (pattern B) were less able to be associated to the amoeba surface than pattern A and C viruses (Fig. 8B). To evaluate the incorporation of lineage A, B, or C isolate particles by amoebas, 50 cells were analyzed by TEM at 30 min pi for each isolate. It is important to mention that these results should be analyzed with care, considering that TEM was performed on a given section of biological samples. Nevertheless, these results are worthwhile to be presented since cells infected by APMV, Borely moumouvirus, or Megavirus caiporensis were analyzed under the same conditions. Here, 43 of 50 analyzed amoebas had at least two APMV particles internalized, and some cells incorporated a substantial number of APMV particles, up to 34 (Fig. 8C). As for Borely moumouvirus-inoculated sample, only 4 infected cells were observed, whereas 16 analyzed cells showed Megavirus caiporensis particles internalized (Fig. 8C). Despite those differences in adhesion and entry described for pattern A, B, and C viruses, we were able to observe particle uncoating (stargate opening) for APMV, Borely moumouvirus, and Megavirus caiporensis (Fig. 8D). Taken together, our results suggest that the pattern of fibrils may affect not only particles adhesion but also their incorporation by amoebas. Finally, the viral factories of viruses belonging to patterns A, B, and C were analyzed by TEM, and this revealed that, in spite of differences in fibril abundances and organization, factories from all viruses presented a fibrils acquisition area, suggesting that further studies on these structures are necessary (Fig. 8E).

TABLE 1 Similarity values of the best hits found by the BLASTp tool from amino acid sequences of proteins R135, L829, and L275[a]

	% similarity with protein:		
Virus	**R135**	**L829**	**L725**
APMV	100	100	100
Mimivirus reunion	99.6	99.8	100
Acanthamoeba castellanii mamavirus	99.6	99.8	100
Cotonvirus japonicus	68.4	61.3	66
Borely moumouvirus	68.2	62.9	54.7
Moumouvirus australiensis	68	61.1	55.1
Moumouvirus maliensis	68	62.5	54.7
Moumouvirus monve	68.5	62.2	55.2
Moumouvirus goulette	68.5	61.2	54.3
Acanthamoeba polyphaga moumouvirus	68.5	62.2	55.6
Saudi moumouvirus	68.5	62.3	55.6
Megavirus caiporensis	67.4	59	56.5
Bandra megavirus	68.3	58.5	56.5
Megavirus courdo7	68	58.8	56.5
Megavirus chilensis	68.1	58.3	56.5
Megavirus vitis	68	58.8	56.5
Megavirus Iba	68	59	56.5
Powai Lake megavirus	67.8	60.1	56.5

[a]APMV was used as the reference.

DISCUSSION

The isolation of a new lineage C mimivirus in Brazil highlights the diversity and ubiquity of mimiviruses. The role of fibrils for the adhesion of mimiviruses and their consequent success in invading their hosts to start the infection cycle has been reported previously (4–6). Findings from this current work corroborate these data, as we provided evidence that fibril abundance may be related to the entry of mimivirus into amoebas. It is important to highlight that lineage A includes the majority of isolates described worldwide, and most of the mimiviruses isolated by our team in Brazil are related to this lineage (23, 24). A possible explanation for this is the morphology of the fibrils presented by these viruses, which are found abundantly and homogeneously distributed throughout the viral capsid (Fig. 3 and 5). Additionally, the smaller number of isolates related to lineage B can also be related to the same hypothesis, since these viruses have fewer fibrils, which can decrease their abilities to adhere to their hosts and trigger phagocytosis and the consequent entry.

Additionally, regarding fibril morphology, it is also possible to notice an interspecific pattern of clustering of these structures within lineage C (Fig. 5, third column). The organization of the fibrils in clumps starts at the insertion from the capsid and continues until the tips facing the external environment. However, in some particles, these clumps were in smaller quantities and presented with more contrast in TEM images (Fig. 5, Megavirus botiensis and Megavirus muiraquitaensis), while they were more abundant and seemed to cluster fewer fibrils in others (Fig. 5, Megavirus caiporensis, Megavirus curupirensis, and Megavirus boitataensis). Understanding the reason for these differences within the same clade, searching for molecular differences that may explain them, and understanding the implications of these data are important perspectives to complement the novelties brought by this work.

Regarding their structure, we also showed that the fibril morphotypes were not the result of preparations for EM analysis but were an intrinsic characteristic of each viral particle pattern's specificity (Fig. 4). Moreover, we found some evidence of molecular differences at important proteins for mimivirus fibrils, namely, R135, L829, and L725 (Fig. 7A to C), which may be closely related to the different phenotypes found among the three lineages. It is important to mention that the scarce number of images from lineage B mimivirus viral particles in the literature was a limiting factor in our study, making our findings concerning fibril patterns dependent on the isolates of each

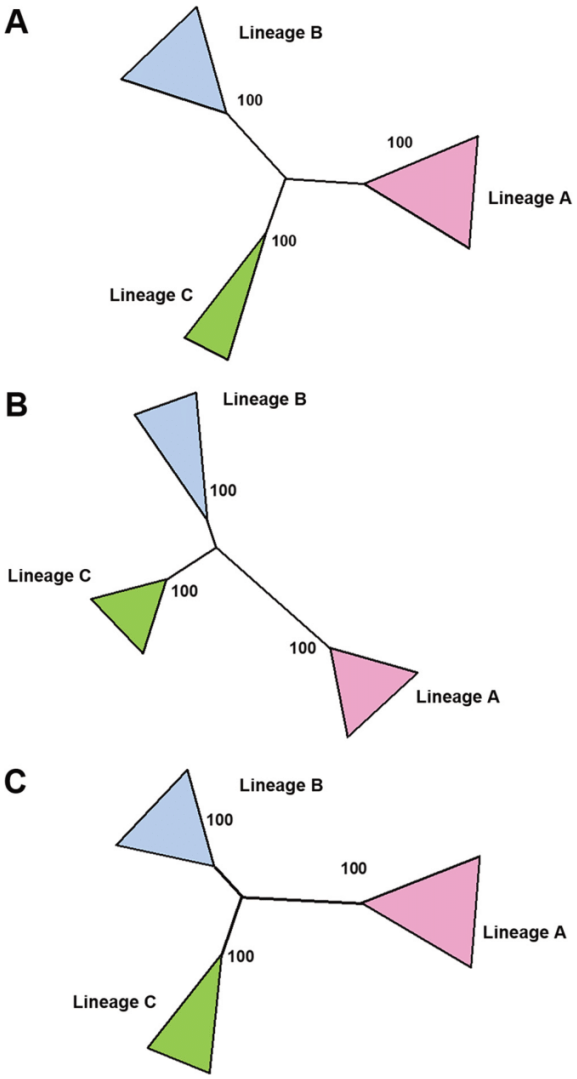

FIG 7 Phylogenetic constructions based on amino acid sequences of proteins associated with mimivirus fibril structures, based on proteins R135 (A), L829 (B), and (C) L275. Our results suggest that those genes are widespread in mimivirus isolates, regardless the lineage.

lineage to which we had access in our database. In addition, although it was clear that the pattern B remained among the isolates of our group, it is essential to attempt to cover future studies with more isolates of lineage B in order to affirm that this pattern can be maintained. Thereby, we propose that fibril morphological characteristics can be considered as indicators of mimivirus relationships to one of the A, B, or C lineages and that they can be complements to molecular traits for the characterization of these viral entities. Understanding the implications of different fibril morphotypes among mimivirus lineages can provide important information about the biology of these viruses and the way they relate to the environment and to their hosts, allowing further unraveling of these features within the family *Mimiviridae*.

MATERIALS AND METHODS

Virus isolation, multiplication, and purification. In August 2017, 10 water samples were collected from Pampulha Lagoon, Belo Horizonte, Brazil. The collection was performed with sterile tubes, and the

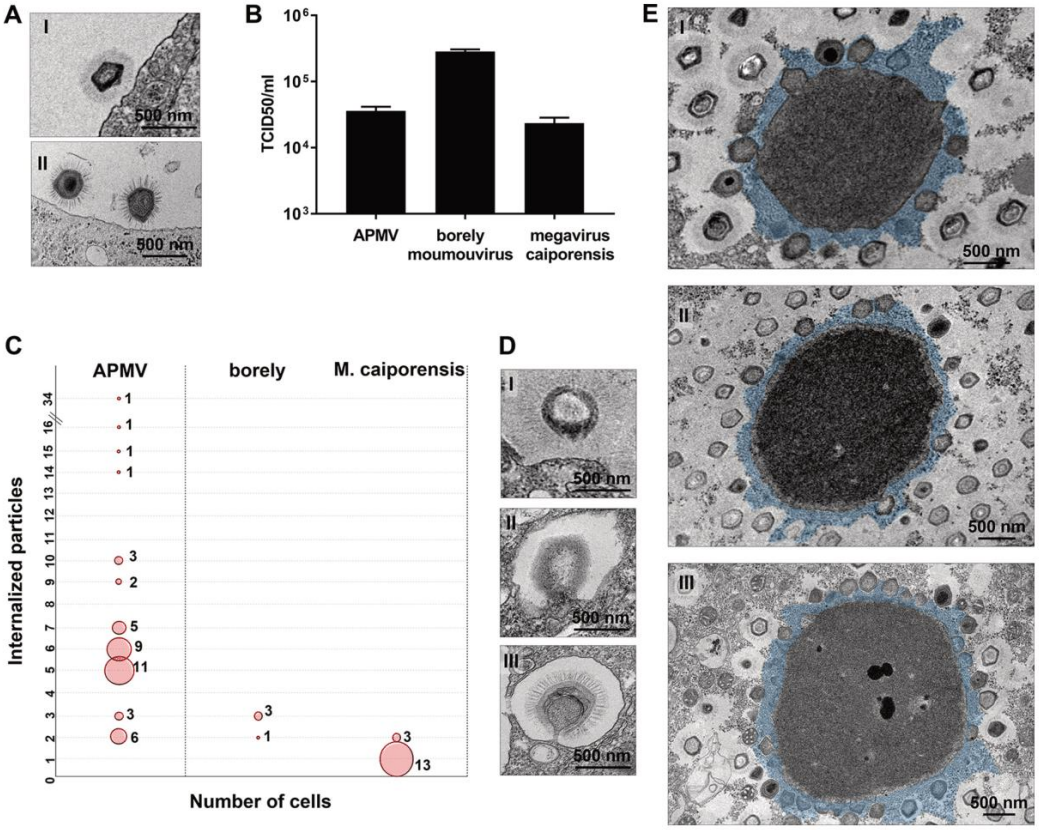

FIG 8 Fibril abundance seems to influence host-amoeba adherence and entry of pattern A, B, and C mimiviruses. (A) TEM images of pattern A (I) and C (II) mimiviruses attached to *Acanthamoeba castellanii* plasma membrane. No images were obtained for pattern B mimivirus. (B) Pattern B mimivirus particles are less incorporated by amoebas than mimivirus pattern A and C particles. Amoebas were infected by pattern A, B, or C at an MOI of 10. Sixty minutes postinfection, the supernatants were collected to verify the number of infectious particles not incorporated by amoebas, and then titers were determined. (C) Graphical representations of the relationship of the number of internalized particles within each amoeba (*y* axis) versus the number of cells (*x* axis) for APMV, Borely moumouvirus, and Megavirus caiporensis, respectively. The red circles indicate the number of cells with a specific number of particles internalized. Fifty TEM images of amoebas for each virus were selected randomly for analysis. (D) Mimivirus particles in denudation after entry into the host cell: APMV (I), Borely moumouvirus (II), and Megavirus caiporensis (III), indicating the continuation of the mimivirus cycle. (E) Fibril acquisition areas of mimiviruses related to lineages A, B, and C. Despite the differences in the patterns of fibrils among the three lineages, all three representatives presented a fibril acquisition area (I, APMV; II, Borely moumouvirus; (III), Megavirus caiporensis).

samples were stored at 5°C until the inoculation process. The samples were submitted to the process of prospecting and isolation of new giant viruses in *Acanthamoeba castellanii* amoebas, as previously described (25). From this collection of samples, we isolated nine amoeba viruses, including Megavirus caiporensis (mimivirus). These new isolates were registered at the Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SisGen), number ABF23CC. After the isolation, the virus was inoculated at a MOI of 0.01 in cell culture roller flasks containing 1.4×10^7 *Acanthamoeba castellanii* cells and 35 mL of peptone-yeast extract-glucose (PYG) medium supplemented with penicillin (100 U/mL; Cellofarm, Brazil), streptomycin (100 μg/mL; Sigma-Aldrich, Burlington, MA, USA), and amphotericin B (0.25 μg/mL; Cultilab, Brazil). The cells were incubated at 30°C under slow rotation (0.2 rpm) on a roller. After the observation of cytopathic effect caused by viral infection (i.e., rounding cells and cellular lysis), the flask contents were collected. This content was subjected to freezing and thawing three times, to lyse the cells that remained intact and release viral particles. Then, it was ultracentrifuged (36,000 $\times$ *g*) in a 22% sucrose cushion for 50 min. The pellet containing purified viral particles was suspended in phosphate-buffered saline (PBS), and the viral titers were obtained by the endpoint method (26).

Electron microscopy. *Acanthamoeba castellanii* cultures infected with different mimivirus strains (belonging to the lineages A, B, and C) and the respective isolated and purified mimiviruses were analyzed by SEM and TEM. Experiments and analyses were performed in the Center of Microscopy at the Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil (http://www.microscopia.ufmg.br).

For SEM assays, samples were fixed by immersion in a solution containing glutaraldehyde (2.5%) in 0.1 M sodium cacodylate buffer (pH 7.2) for 2 h. Following this, postfixation was performed for each sample with 2% osmium tetroxide (OsO_4) for 2 h at room temperature. Fixed samples were dehydrated using a growing series of ethanol dilutions (30%, 50%, 70%, 95%, and 100%) for 10 min in each step. Then, samples were dried with CO_2 at critical point using CPD 030 equipment (Bal-Tec, Liechtenstein). Next, samples were supported in aluminum stubs and metalized with a thin layer (5 nm) of gold particles using MED 020 equipment (Bal-Tec, Liechtenstein). Samples were observed in a FEG-Quanta 200 FEI microscope (FEI Co., Eindhoven, Netherlands) at 15 to 20 kV.

For TEM assays, samples were fixed by immersion in a solution containing glutaraldehyde (2.5%) in 0.1 M sodium phosphate buffer (pH 7.2) for 2 h. After fixation, postfixation was performed with a solution of 1% osmium tetroxide in sodium cacodylate buffer (0.1 M, pH 7.2) for 1 h followed by en bloc counterstaining with uranyl acetate (2% uranyl acetate in deionized water). Samples were gradually dehydrated by immersion in 70%, 80%, and 90% ethanol once for 15 min each and twice in 100% ethanol for 15 min. Next, samples were embedded in Epon resin. Ultrathin sections were obtained using an ultramicrotome with diamond knives (Leica Microsystems), and these sections had a thickness of 70 nm and were placed on a 200 mesh copper screen. The screens were counterstained with Reynold's lead citrate solution for 10 min. Images were obtained using a Tecnai G2-12-SpiritBiotwin FEI electron microscope (FEI Co., Eindhoven, Netherlands) at an acceleration voltage of 120 kV using a charge-coupled-device camera. In order to obtain the Megavirus caiporensis particle medium dimensions, the ImageJ software (version v1.53k, National Institutes of Health) was used to measure seven different particles during the acquisition of images. The measures were used to calculate the medium sizes of the particle and the fibrils.

Sequencing, assembly, and annotation. The samples containing purified virus were sequenced using the Illumina MiSeq system, with a paired-end library and an Illumina DNA Prep kit (Illumina Inc., San Diego, CA, USA). The FastQC program was used to quality control of the obtained reads, which were trimmed using the Trimmomatic tool (27). For genome *de novo* assembly, we used Spades 3.12 with default parameters (28), and the obtained scaffolds were ordered based on a reference genome using MeDuSa online (29). The reference genome used was that of Powai Lake megavirus, obtained from the NCBI database (GenBank accession number KU877344.1). The GeneMarkS tool (30) was used to predict ORFs, considering only proteins that were larger than 50 amino acids. Additionally, the predictions of tRNA coding sequences were performed using Aragorn (31). The predicted ORFs were annotated using BLASTp (expect threshold, 10^{-3}) against the NCBI non-redundant protein sequences (nr) database.

Phylogeny analysis. Phylogenetic trees were constructed employing IQtree software (version 1.6.12) using the maximum-likelihood statistic method, with 1,000 bootstrap replicates as branch support (32). The mimivirus subfamily B DNA polymerase and the gene sequences for proteins R135, L829, and L725 were used. The data sets containing the sequences used for alignments were prepared using BLASTp (expected threshold, 10^{-3}) against the NCBI non-redundant protein sequences (nr) database. The sequences selected were aligned by the Muscle algorithm executed by MEGAX (33, 34). The best-fit substitution models were selected by the ModelFinder algorithm implemented in IQtree (35), and the visualization and editing of the phylogenetic trees were carried out with MEGAX software and iTOL (34, 36).

Fibril morphology and density analysis. In order to analyze differences between the fibrils' morphology and density, we developed some specific protocols. To highlight the fibrils and their morphological differences, TEM images of three mimiviruses were used: APMV (representing lineage A), Borely moumouvirus (representing lineage B), and Megavirus caiporensis (the isolate described here, representing lineage C). First, fibrils were digitally highlighted with colors, emphasizing their peculiarities. For this, fibrils were marked with the aid of the "Marker" or "Watercolor" drawing tools of the Paint 3D (version 2021) program (Microsoft Corporation), with thickness settings between 2 and 4 pixels. Single colors were used for each mimivirus, with a maximum opacity of 35%, to highlight the fibrils with colors without losing their original shape.

For the density analysis, a protocol commonly used to measure the electrophoretic gel band densities was adapted (37). First, 70 TEM images for each one of the three mimiviruses representing lineages A, B, and C were individually highlighted with the tool "Scribble" of Microsoft PowerPoint version 2021 (Microsoft Corporation) very carefully (Fig. 3C, image I). Then, the background image was deleted, leaving only the outlines of the fibrils that were marked (Fig. 3C, image II). Those images were than used in the program ImageJ (version v1.53k, National Institutes of Health) to read the contrast and estimate the fibril densities for the 6 sides of each particle (previously selected, with many rectangles of same size), until the entire area presenting fibrils was fully covered by them (Fig. 3C, image III). After the selection of areas to be measured, the tool plot lanes were used to estimate the contrast from generated graphics (Fig. 3C, image IV). Each rectangle generated a graph with values related to the pixels detected in the analyzed images. These values were then used to interpret the mean contrast values for each of the six sides of the particles, and a total mean value was calculated, considering the standard deviation. The mean values were then normalized to be plotted in graph form (Fig. 3C, image V).

Evaluation of the relationship between fibrils and entry at the host cell. To assess whether the number of fibrils was associated with the success of mimiviruses to adhere and consequently penetrate the host amoebas, we performed *A. castellanii* amoeba infection assays. The same representatives of lineages A, B, and C used previously were used in this experiment. Cell culture flasks (25 cm^2; Thermo Fisher Scientific, USA) containing 1×10^6 *A. castellanii* amoebas were infected with APMV, Borely moumouvirus, or Megavirus caiporensis at an MOI of 10. After infection, 30 min of adsorption was carried out at 30°C, with the flasks being gently shaken every 10 min to ensure that the inoculum covered the entire adhered cell monolayer. After adsorption, the inoculum was removed and the monolayer was

washed 2 times with PBS, followed by the addition of 5 mL PYG medium to loosen the monolayer and collect the contents of the flasks. We proceeded with sample preparation for TEM analysis as described above. During imaging, we randomly selected 50 amoebas for each of the samples and counted the number of particles found inside each of the cells, to later analyze and compare the results.

We also tested the inoculums of infections performed in 96-well plates containing 4×10^4 amoebas per well, with inoculums at an MOI of 10 using APMV, Borely moumouvirus, and Megavirus caiporensis. After 1 h of adsorption, the inoculums were collected and the viral titers were obtained by the endpoint method (26).

Data availability. The genome of Megavirus caiporensis is available at GenBank under accession number OP925046.

ACKNOWLEDGMENTS

We are grateful to our colleagues from Laboratório de Vírus and Microscopy Center, Universidade Federal de Minas Gerais, and Laboratório de Virologia, Universidade Estadual Paulista for their excellent technical support. In addition, we acknowledge the financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do estado de Minas Gerais (FAPEMIG), Ministério da Ciência, Tecnologia e Inovações (MCTI), and Pro-Reitorias de Pesquisa e Pós-Graduação of UFMG. J.S.A. and J.P.A. are CNPq researchers.

REFERENCES

1. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A giant virus in amoebae. Science 299:2033. https://doi.org/10.1126/science.1081867.
2. Abrahão JS, Dornas FP, Silva LCF, Almeida GM, Boratto PVM, Colson P, La Scola B, Kroon EG. 2014. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. Virol J 11:120. https://doi.org/10.1186/1743-422X-11-120.
3. Villalta A, Schmitt A, Estrozi LF, Quemin ERJ, Alempic J-M, Lartigue A, Pražák V, Belmudes L, Vasishtan D, Colmant AMG, Honoré FA, Couté Y, Grünewald K, Abergel C. 2022. The giant mimivirus 1.2 Mb genome is elegantly organized into a 30 nm helical protein shield. bioRxiv. https://doi.org/10.1101/2022.02.17.480895.
4. Klose T, Herbst D, Zhu H, Max J, Kenttämaa H, Rossmann M. 2015. A mimivirus enzyme that participates in viral entry. Structure 23:1058–1065. https://doi.org/10.1016/j.str.2015.03.023.
5. Rodrigues RAL, dos Santos Silva LK, Dornas FP, de Oliveira DB, Magalhães TFF, Santos DA, Costa AO, de Macêdo Farias L, Magalhães PP, Bonjardim CA, Kroon EG, La Scola B, Cortines JR, Abrahão JS. 2015. Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. J Virol 89:11812–11819. https://doi.org/10.1128/JVI.01976-15.
6. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG. 2009. Structural studies of the giant mimivirus. PLoS Biol 7:e92. https://doi.org/10.1371/journal.pbio.1000092.
7. Boyer M, Azza S, Barrassi L, Klose T, Campocasso A, Pagnier I, Fournous G, Borg A, Robert C, Zhang X, Desnues C, Henrissat B, Rossmann MG, La Scola B, Raoult D. 2011. Mimivirus shows dramatic genome reduction after intraamoebal culture. Proc Natl Acad Sci U S A 108:10296–10301. https://doi.org/10.1073/pnas.1101118108.
8. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH. 2020. Global organization and proposed megataxonomy of the virus world. Microbiol Mol Biol Rev 84:e00061-19. https://doi.org/10.1128/MMBR.00061-19.
9. Palermo CN, Fulthorpe RR, Saati R, Short SM. 2019. Metagenomic analysis of virus diversity and relative abundance in a eutrophic freshwater harbour. Viruses 11:792. https://doi.org/10.3390/v11090792.
10. Boratto PVM, Serafim MSM, Witt ASA, Crispim APC, de Azevedo BL, de Souza GAP, de Aquino ILM, Machado TB, Queiroz VF, Rodrigues RAL, Bergier I, Cortines JR, de Farias ST, dos Santos RN, Campos FS, Franco AC, Abrahão JS. 2022. A brief history of giant viruses' studies in Brazilian biomes. Viruses 14:191. https://doi.org/10.3390/v14020191.
11. Colson P, de Lamballerie X, Fournous G, Raoult D. 2012. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. Intervirology 55:321–332. https://doi.org/10.1159/000336562.
12. Campos RK, Boratto PV, Assis FL, Aguiar ERGR, Silva LCF, Albarnaz JD, Dornas FP, Trindade GS, Ferreira PP, Marques JT, Robert C, Raoult D, Kroon EG, La Scola B, Abrahão JS. 2014. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virol J 11:95. https://doi.org/10.1186/1743-422X-11-95.
13. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A 108:17486–17491. https://doi.org/10.1073/pnas.1110889108.
14. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, Azza S, Klose T, Wong J, Rossmann MG, La Scola B, Raoult D, Koonin EV. 2012. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol Evol 4:1324–1330. https://doi.org/10.1093/gbe/evs109.
15. Andrade A, Rodrigues RAL, Oliveira GP, Andrade KR, Bonjardim CA, La Scola B, Kroon EG, Abrahão JS. 2017. Filling knowledge gaps for mimivirus entry, uncoating, and morphogenesis. J Virol 91:e01335-17. https://doi.org/10.1128/JVI.01335-17.
16. Andrade KR, Boratto PPVM, Rodrigues FP, Silva LCF, Dornas FP, Pilotto MR, La Scola B, Almeida GMF, Kroon EG, Abrahão JS. 2015. Oysters as hot spots for mimivirus isolation. Arch Virol 160:477–482. https://doi.org/10.1007/s00705-014-2257-2.
17. Xiao C, Chipman PR, Battisti AJ, Bowman VD, Renesto P, Raoult D, Rossmann MG. 2005. Cryo-electron microscopy of the giant mimivirus. J Mol Biol 353:493–496. https://doi.org/10.1016/j.jmb.2005.08.060.
18. Akashi M, Takemura M. 2019. Co-isolation and characterization of two pandoraviruses and a mimivirus from a riverbank in Japan. Viruses 11:1123. https://doi.org/10.3390/v11121123.
19. Yoosuf N, Pagnier I, Fournous G, Robert C, La Scola B, Raoult D, Colson P. 2014. Complete genome sequence of Courdo11 virus, a member of the family Mimiviridae. Virus Genes 48:218–223. https://doi.org/10.1007/s11262-013-1016-x.
20. Saadi H, Pagnier I, Colson P, Cherif JK, Beji M, Boughalmi M, Azza S, Armstrong N, Robert C, Fournous G, La Scola B, Raoult D. 2013. First isolation of Mimivirus in a patient with pneumonia. Clin Infect Dis 57:e127–e134. https://doi.org/10.1093/cid/cit354.
21. Saadi H, Reteno D-GI, Colson P, Aherfi S, Minodier P, Pagnier I, Raoult D, La Scola B. 2013. Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia. Intervirology 56:424–429. https://doi.org/10.1159/000354564.
22. Gaia M, Benamar S, Boughalmi M, Pagnier I, Croce O, Colson P, Raoult D, La Scola B. 2014. Zamilon, a novel virophage with Mimiviridae host specificity. PLoS One 9:e94923. https://doi.org/10.1371/journal.pone.0094923.
23. Dornas FP, Khalil JYB, Pagnier I, Raoult D, Abrahão J, La Scola B. 2015. Isolation of new Brazilian giant viruses from environmental samples using a panel of protozoa. Front Microbiol 6:1086. https://doi.org/10.3389/fmicb.2015.01086.
24. Andrade A, Arantes TS, Rodrigues RAL, Machado TB, Dornas FP, Landell MF, Furst C, Borges LGA, Dutra LAL, Almeida G, Trindade GdS, Bergier I,

Abrahão W, Borges IA, Cortines JR, de Oliveira DB, Kroon EG, Abrahão JS. 2018. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. Virol J 15:22. https://doi.org/10.1186/s12985-018-0930-x.
25. Machado TB, de Aquino ILM, Abrahão JS. 2022. Isolation of giant viruses of Acanthamoeba castellanii. Curr Protoc 2:e455. https://doi.org/10.1002/cpz1.455.
26. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. Am J Epidemiol 27:493–497. https://doi.org/10.1093/oxfordjournals.aje.a118408.
27. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
28. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102.
29. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–2451. https://doi.org/10.1093/bioinformatics/btv171.
30. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–454. https://doi.org/10.1093/nar/gki487.
31. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
32. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
33. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
34. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.
35. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285.
36. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301.
37. Gallagher SR. 2014. Digital image processing and analysis with ImageJ. Curr Protoc Essential Lab Tech 9:A.3C.1–A.3C.29. https://doi.org/10.1002/9780470089941.eta03cs9.

Minireview

Surface fibrils on the particles of nucleocytoviruses: A review

Isabella Luiza Martins de Aquino, Matheus Gomes Barcelos, Talita Bastos Machado, Mateus Sá Magalhães Serafim and Jônatas Santos Abrahão

Laboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil
Corresponding author: Jônatas Santos Abrahão. Email: jonatas.abrahao@gmail.com

Impact Statement

Considering the scenario of multiple origins of viruses along the viral evolution, many peculiar structural features have been reported and described for nucleocytoviruses. In this context, capsids of some of these so-called giant viruses (GV) are highlighted by surface fibrils, which are responsible for adhesion and interaction with their host membranes, such as those in amoeba. Knowing the singularities of these structures promotes the deepening of fundamental knowledge on the biology of these members within the virosphere, ultimately shedding light on their adaptive convergence in evolution, virus–host interactions, and their biological cycle, as well as possibilities for their proteins and viral assembly.

Abstract

The capsid has a central role in viruses' life cycle. Although one of its major functions is to protect the viral genome, the capsid may be composed of elements that, at some point, promote interaction with host cells and trigger infection. Considering the scenario of multiple origins of viruses along the viral evolution, a substantial number of capsid shapes, sizes, and symmetries have been described. In this context, capsids of giant viruses (GV) that infect protists have drawn the attention of the scientific community, especially in the last 20 years, specifically for having bacterial-like dimensions with hundreds of different proteins and exclusive features. For instance, the surface fibrils present on the mimivirus capsid are one of the most intriguing features of the known virosphere. They are 150-nm-long structures attached to a 450-nm capsid, resulting in a particle with a hairy appearance. Surface fibrils have also been described in the capsids of other nucleocytoviruses, although they may differ substantially among them. In this mini review for non-experts, we compile the most important available information on surface fibrils of nucleocytoviruses, discussing their putative functions, composition, length, organization, and origins.

Keywords: Adhesion, fibrils, giant viruses, virus structure

Experimental Biology and Medicine **2023; 248: 2045–2052. DOI: 10.1177/15353702231208410**

Introduction

A glimmer of microbial evolution: giants among viruses

The term "giant viruses" has been used to designate a putative monophyletic group of viruses belonging to the phylum Nucleocytoviricota, infecting uni- and pluricellular, hetero- and autotrophic organisms, from protists to animals.[1] Some authors consider bona-fide giant viruses those with capsids larger than 500 nm, which are easily visible by optical microscopy. However, the term "giant viruses" has also been used to refer to smaller nucleocytoviruses, with particles ranging from 150 to 500 nm. Giant viruses are ubiquitous and have already been isolated from different countries,[2–5] from different environmental[6,7] and clinical samples.[8] The first virus to be called a giant virus was Paramecium bursaria chlorella virus 1 (PBCV-1), which infects algae and which has particles up to approximately 190 nm in diameter,[9] and a dsDNA genome of approximately 330 kb.[10]

Although nucleocytoviruses include historically important representatives, such as poxviruses,[11] the discovery of mimivirus in 2003[12] highlighted the remarkable structural complexity of the virions of this group.

Among mimiviruses, the Acanthamoeba polyphaga mimivirus (APMV)[12] isolate was the first amoebae-associated GV to be described. APMV was isolated from water samples collected from a hospital cooling tower during a pneumonia outbreak in Bradford, England. After analyzing these samples by Gram stain, researchers noticed small purple-stained dots inside amoebae, similar to gram-positive bacteria. However, after the failure of several techniques used in the identification and characterization of bacteria, questions about the nature of these organisms remained for years. After almost a decade, new techniques were used to study the mysterious microorganism, such as genome sequencing and electron microscopy, which led to the surprising discovery of an actual virus, and not bacteria.[12] This virus attracted the attention of the scientific community due to

ISSN 1535-3702

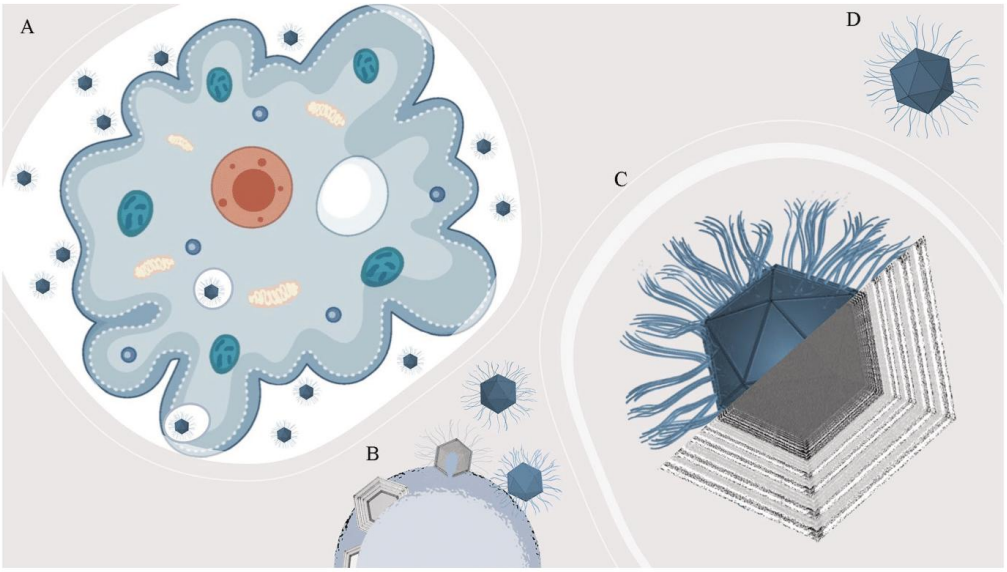

Figure 1. Fibrils' role for giant viruses. (A) Fibrils (~150 nm) are important to the adhesion of mimiviruses onto amoebas, such as APMV. (B) During the morphogenesis, fibrils seem to be acquired in the fibril acquisition area, located at the periphery of the viral factory, concomitantly with the genome acquisition. (C) Fibrils' peptidoglycan-like structures exhibit successive rings of density under electronic microscopy. (D) Schematic representation of a mature APMV particle. Amoeba image was generated from free vectors available at Vecteezy (https://www.vecteezy.com).

the large size of its particle and its genome. The APMV particle is approximately 750 nm in diameter, and its genome reaches the million mark, about 1.2 megabases (Mb). Since then, several other amoeba GV groups have been discovered, such as marseillevirus,[13] pandoravirus,[14] pithovirus,[15] and cedratvirus,[16] among others.

There is great diversity within giant viruses, considering their morphological characteristics. In general, the capsids of giant viruses do not present an external envelope. Instead, the capsids surround an inner lipid sac, which contains the viral genome and proteins related to the early phases of the replication cycle. Particles can have icosahedral symmetry (e.g. marseillevirus[13] and faustovirus[17]), or pseudoicosahedral symmetry, as in mimiviruses, due to the presence of a stargate, that is, a vertex at the capsid that allows for DNA release.[18] They may also be oval (e.g. pithovirus[15] and pandoravirus[14]) or even round-shaped (e.g. mollivirus[4]). For instance, one of the largest viruses ever described, Tupanvirus, has a capsid attached to a tail variable in size, allowing the particle to reach up to 2.3 µm.[19] In some GV, we can find some structures decorating their capsids, such as the spherical-headed spikes that cover medusaviruses' surfaces and mimiviruses' fibrils.[20]

Viral fibrils: an intriguing structural feature among giants

As aforementioned, mimiviruses exhibit particles with a capsid of approximately 750 nm, being 450 nm in diameter and covered by a dense layer of fibrils (~150 nm),[21] which were suggested to resemble gram-positive bacteria.[12] This intriguing misunderstanding would be feasible due to its structural fibrils' composition, which is morphologically unique among these viruses,[22] and not being fully elucidated yet,[23] with all that is currently known being related to mimiviruses. The fibrils are embedded in a dense layer of peptido glycan-like structures[24] that stain crystal violet in a Gram stain, as those peptidoglycans that are present in the walls of gram-positive bacteria.[25] Important to the adhesion onto the amoeba surface (Figure 1(A)), as for APMV, fibrils are often found with one of their ends free while the other end is attached to the viral capsid.[26] Moreover, scanning electron microscopy (SEM) and transmission electron microscopy (TEM) analyses have shown different densities of fibrils on GV capsid surfaces, which are simultaneously acquired with the genome acquisition during morphogenesis in the fibril acquisition area, located at the periphery of the viral factories[27] (Figure 1(B)). In addition, fibrils' peptidoglycan-like structures exhibit successive so-called rings of density (Figure 1(C)), which supports their key role in viral entry,[26] as exemplified for APMV (Figure 1(D)).

Many GV also present fibrils (Figure 2) as APMV, such as: (1) tupanviruses, with a capsid of approximately 450 nm in diameter and a tail attached to it, both covered in fibrils, resulting in particles varying from 1.2 up to 2.3 µm in length;[19] (2) *Cotonvirus japonicus*,[28] a mimivirus isolate, with approximately 400 nm in diameter and also surrounded by surface fibrils of approximately 100 nm; (3) Marseillevirus,[13] with approximately 250 nm of diameter, and 12 nm fibrils surrounding its particles' surface; (4) PBCV-1, an algae

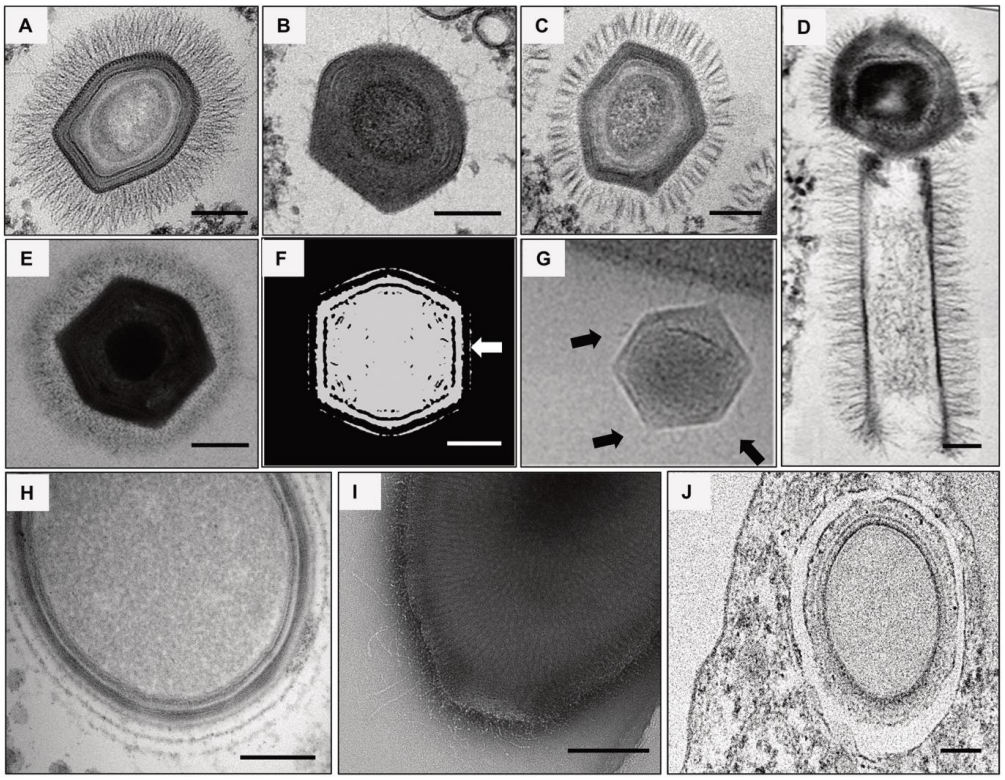

Figure 2. Giant viruses that possess surface fibrils. Electron microscopy images and shaded-surface representation of different GV isolates and their fibrils. (A–C) Lineage A, B, and C mimiviruses: mimivirus puntaullmanensis, moumouvirus crenensis, and megavirus caiporensis, respectively. (D) Tupanvirus particle, with its tail, a trademark of this group (adapted from DOI: 10.1038/s41467-018-03168-1). (E) *Cotonvirus japonicus* (adapted from DOI: 10.1128/JVI.00919-21). (F) Marseillevirus shaded representation illustrating the globular portions of its fibrils indicated by the white arrow (adapted from DOI: 10.1038/s41467-018-03168-1). (G) PBCV-1 (adapted from DOI: 10.1073/pnas.1107847108) fibrils are pointed by the black arrows. (H) *Mollivirus sibericum* (credits: provided by Dr Chantal Abergel, IGS UMR7256 CNRS-AMU). (I) *Cedratvirus pambiensis*. (J) Orpheovirus IHUMI-LCC2. Scale bars: 150 nm.

infecting virus with approximately 190 nm in length and small fibrils around the capsid, of nearly 19 nm;[9] (5) orpheovirus, with particles larger in length than APMV (up to 1.1 µm), but presenting smaller fibrils when compared to these mimiviruses, as well as two tegument layers between the fibrils and the inner membrane;[29] (6) *Mollivirus sibericum*,[4] an almost spherical GV with 500–600 nm in diameter and surrounded by a fibrillar tegument as a mesh of fibrils; (7) cedratviruses, around 500 nm up to 1 µm in length,[16,30] whose viral factory can be divided in two, one being associated with its fibrils acquisition; and finally (8) *Yasminevirus*,[31] the first *Klosneuvirus* to be isolated and that presents a particle with approximately 330 nm covered by a thin layer of fibrils.

For instance, Notaro *et al.* demonstrated that fibrils are glycosylated with two different polysaccharides in their composition, whose structures have been elucidated by nuclear magnetic resonance (NMR) analysis, one being L-rhamnose and *N*-acetylglucosamine (GlcNAc) arranged in a linear repeating pattern, and the other as a branched repeating unit. Chemical analysis of the monosaccharide composition as acetylated *O*-methyl glycosides confirmed L-rhamnose (L-Rha), GlcNAc, and 2-*O*-methyl-4*N*-acetyl-viosamine (2OMeVio4NAc), including minor amounts of non-methylated Vio4NAc, mannose, and glucose. In addition, glycan-like, or containing, molecules were also a fraction (13.7%) of the structure, showing the influence of the protease digestion protocols in removing protein fractions of the fibrils.[32] The glycan composition in fibrils would be related to their binding to host cells, mediating their interactions with hosts' surface glycans (e.g. mannose and GlcNAc), triggering phagocytosis,[24] whereas fibrils' peptide fractions are suggested to be putatively more associated with the 2-glucose-methanol-choline (GMC) oxireductase's structural composition and assembly.[23]

Notaro *et al.* also showed rare amino-acid sugars to be synthesized from different GV of the subfamily Megamimivirinae, which are absent from their hosts and usually found only in bacteria: (1) D-viosamine and rhamnose in

the A-clade; (2) D-Qui2NAc and D-Fuc2Nac in the B-clade; (3) L-RhaNAc and L-Qui2Nac in the C-clade; and (4) bacillosamine (D-diNacBac) observed only in a few GV so far (*Moumouvirus australiensis*, *Cotonvirus japonicus*, and tupanviruses). Thus, this varied set of glycans decorating capsids could also support the discussion that these structures could mimic bacteria that amoeba feed on, facilitating competition among different organisms and different GV competing for the same host. In addition, differences in sugar compositions and biosynthetic pathways for nucleotide sugars and glycosyltransferases were shown to be related to complex gene clusters (e.g. 6 and up to 33 genes), which are involved in the glycosylation of fibril layers among the subfamily Megamimivirinae and were also shown to be clade-specific.[33]

Harboring giants' structural repertoire: genes associated with fibrils

Initially, it was speculated that the mimiviruses' hairy-like appearance could be related to the large number of open reading frame exhibiting the characteristic collagen triple-helix repeat. Thus, fibrils could be linked to collagen and glycosylated proteins, consisting of cross-linked glycosylated collagen.[34] However, the fibrils are resistant to collagenases, even after pre-treatment with lysozyme, suggesting that the fibril surfaces are not collagen-linked.[35] For instance, the L829 and L725 genes, which codify proteins with unknown functions, and the R135 gene, which codifies a putative GMC-oxidoreductase, were also associated with mimivirus fibrils in previous works.[36,37] Boyer *et al*. obtained an artificial mimivirus strain called M4 after several blind passages of APMV in axenic amoebae culture. M4 has a reduction of ~200,000 bp in its genome, and its particles have a bald appearance with fewer fibrils. After analysis of purified viral fibers by 2D gel electrophoresis coupled with matrix-assisted laser desorption/ionization mass spectrometry, it was reported that there was an absence of R135, L829, and L725 proteins in M4 particles when compared to the original mimivirus M1. They also compared the protein glycosylation patterns in the M1 and M4 viruses and observed that L829 and R135 proteins were glycosylated only in M1.[36] Furthermore, L725, which is encoded by an ORFan gene, that is, genes with no detectable sequence similarities in databases,[38] was suggested to be associated with fibrils by RNA silencing experiments.[39]

In this sense, the R135 protein, APMV's putative GMC-oxidoreductase, was pointed to be part of the fibrils composition and of the helical protein shell that encases mimivirus genomic material by cryo-electron microscopy, cryo-electron tomography, and proteomics,[32,37] with the only difference between the fibrils and the genomic fiber composition being the presence of a Cys-Pro rich N-terminal domain only in the fibrils. In addition, Aquino *et al.* also discussed those genes in contrast to their previous functionalities or relationships with fibrils, as isolates with fewer fibrils (e.g. Borely moumouvirus) are morphologically similar to mimivirus M4 and present R135 and L829 genes in their genomes.[40] Interestingly, these three predicted proteins can also be found only in some GV that have fibrils when analyzing search results using the BLASTp tool[41] (Table 1) from the National Center for Biotechnology Information (NCBI). Herein, the best hits were found for all the proteins among mimiviruses from lineages A, B, and C, and for *Cotonvirus japonicus*, even though lineage B mimiviruses are almost fiberless. As for tupanviruses, a best hit for R135 was not found, and only a best hit for L829 was found for molliviruses, marseillevirus, and cedratvirus. No best hits were found for any of the proteins in PBCV-1 or yasminevirus, which also raises questions about the relationship of these proteins with GV's fibrils. Furthermore, it is also important to highlight that the best hits were also found for R135 and L829 in viruses that do not have surface fibrils described as part of their particles, like pandoravirus and Pithovirus.[40]

Viral attachment: depicting the role of fibrils

Considered intriguing structural features of viruses' morphology until then, fibrils stood out in early studies, and their function remained unknown for more than 10 years after the discovery of APMV. As mentioned, it was described that they play an important role in triggering the host amoeba's phagocytosis by promoting the adhesion of viral particles to the cell surface, mediated by glycans.[24] Interestingly, in the presence of high concentrations of certain glycans (e.g. GlcNAc), interaction of the viral particles with other molecules would be prevented, due to the fibrils being saturated by a given carbohydrate. In addition, a smaller number of fibrils does not alter APMV replication but decreases its attachment to *Acanthamoeba castellanii* cells.[24]

Rodrigues *et al*. treated APMV particles with different enzymatic conditions, such as lysozyme and lysozyme followed by bromelain and proteinase K, both in comparison to APMV and the M4 mimivirus. Interestingly, the presence of more fibrils leads to an increase in viral attachment to the amoebal surface when mediated by mannose and GlcNAc, no changes in viral titers were observed in the presence of glucose or *N*-acetylgalactosamine (GalNAc) at any of the concentrations assessed. The presence of mannose and GlcNAc at different concentrations (>50 and >25 μg/mL, respectively) reduced the viral titer up to 1000-fold. In addition, APMV fibrils also attach differentially to distinct organisms or structures by their glycoside interactions, such as *Aedes* sp. legs (i.e. chitin), *Aspergillus fumigatus* (i.e. mannose and chitin), *Staphylococcus aureus* and *Escherichia coli* (i.e. peptidoglycan), resulting in 18-fold, 7.32-fold, and no differences of the viral particles compared with the control, respectively. Finally, particles were saturated with chitin and peptidoglycan, and the presence of both polymers interfered with APMV adhesion to *A. castellanii* cells, with up to a 100-fold reduction in viral titer at concentrations >25 and >75 μg/mL, respectively.[24]

These results could suggest that the peptidoglycan-like fraction of fibrils would also be restricted to interactions with *Acanthamoeba* cellular plasma membranes,[42] which are rich in lipophosphoglycan,[43] as well as potentially guaranteeing their host specificity. Furthermore, other functions have been suggested for fibrils, such as: (1) optimizing phagocytosis by expanding particle size;[44] (2) stimulating phagocytosis in amoeba by partially mimicking bacteria with peptidoglycan-like compounds, which are "food" of amoebae;[45] (3) acting as a natural decoy for hosts; and (4) increasing resistance under

Table 1. BLASTp best hits for the three genes associated with fibrils: L725, L829, and R135.

Virus	Gene								
	L725			L829			R135		
	Accession number	Size (aa)	Function	Accession number	Size (aa)	Function	Accession number	Size (aa)	Function
APMV	YP_003987254.1	224	Hypothetical protein	YP_003987362.1	433	Hypothetical protein	YP_003986627.1	720	GMC oxidoreductase
Borely moumouvirus	QID05925.1	223	Hypothetical protein	QID05934.1	395	Hypothetical protein	QID06495.1	661	Choline dehydrogenase-like protein
Megavirus caiporensis	WBF70980.1	225	Hypothetical protein	WBF70952.1	663	Hypothetical protein	WBF70396.1	663	Hypothetical protein
Cotonvirus japonicus	BCS82919.1	225	Hypothetical protein	BCS83575.1	387	Hypothetical protein	BCS83533.1	672	GMC oxidoreductase
Tupanvirus deep ocean	QKU34617.1	238	Hypothetical protein	QKU34326.1	432	Hypothetical protein	–	–	–
Tupanvirus soda lake	QKU35955.1	238	Hypothetical protein	QKU35587.1	432	Hypothetical protein	–	–	–
Mollivirus sibericum	–	–	–	YP_009165460.1	348	Hypothetical protein	–	–	–
Mollivirus kamchatka	–	–	–	QHN71453.1	418	Hypothetical protein	–	–	–
Orpheovirus	–	–	–	YP_009449253.1	507	Hypothetical protein	YP_009448335.1	546	GMC oxidoreductase
Cedratvirus kamchatka	–	–	–	QIN54245.1	378	Hypothetical protein	–	–	–
Marseillevirus marseillevirus	–	–	–	YP_003406851.1	400	Hypothetical protein	–	–	–
Yasminevirus	–	–	–	–	–	–	–	–	–
PBCV-1	–	–	–	–	–	–	–	–	–

aa: amino acid; APMV: Acanthamoeba polyphaga mimivirus; BLAST: basic local alignment search tool; BLASTp: Protein BLAST; GMC: glucose-methanol-choline; PBCV-1: Paramecium bursaria chlorella virus 1.

adverse conditions, but more studies remain necessary[26] to better understand those propositions.

Uniqueness of an intriguing structure: diversity in surface fibrils pattern

A first glimpse of fibril-like structures was observed with SEM and TEM methods for chloroviruses termed "fibers" at the time.[46] Subsequent structural[26] studies of amoeba GV, such as mimiviruses, helped in the discovery of other GV[22] and their fibrils. As fibrils of mimiviruses are polysaccharide-containing structures likely built from their glycosylation machinery, which decorates the capsid as surrounding structures,[11] one could argue what the disposition, organization, and composition of different GV's fibrils would impact their biological cycle and even host specificities. Cultivation-independent approaches (e.g. metagenomics) have enhanced the discovery of new genome sequences of potentially new GV, with the potential for structural and functional diversity of fibrils. Although there are some similarities in how these viruses enter host cells (e.g. fibrils triggering phagocytosis), much is yet to be found about GV and their fibrils, as they can be found nearly anywhere on Earth, and only a small fraction of GV's genomes have been discovered so far.[11,47] In this sense, experimental characterization and validation, as demonstrated by Aquino *et al.*, help unravel these differences among different GV's fibrils. Authors showed that mimiviruses from lineages A (e.g. APMV), B (e.g. Borely moumouvirus), and C (e.g. Megavirus caiporensis) present differences regarding their surface fibrils organization and disposition. SEM and TEM allowed authors to observe that fibrils are (1) long and abundantly surrounding the capsid; (2) fewer in number and less homogeneously distributed; or (3) similarly abundant to the first, but organized in small groups (i.e. clusters or clumps).[40]

For instance, APMV fibrils are homogeneously distributed, whereas Megavirus caiporensis appears in clumps or clusters. Both were estimated for their relative abundance of surface fibrils, whereas Megavirus caiporensis exhibited a 552-fold relative average contrast increase when compared to Borely moumouvirus, in comparison to a 394-fold increase for APMV. In addition, different combinations of the three mimiviruses mixed with purified particles were also compared by TEM, disregarding a potential SEM preparation influencing fibrils appearance and conformation. In addition, adhesion and entry into *A. castellanii* were also quantitatively assessed at a multiplicity of infection (MOI) of 10, showing that APMV and Megavirus caiporensis particles bind to the host plasma membrane approximately 10 times more than Borely moumouvirus after 1 h post infection. Finally, more APMV-infected cells were observed (43/50), followed by Megavirus caiporensis (16/50), and Borely moumouvirus (4/50), both under the same experimental conditions.[40] These results could suggest that the pattern of fibrils may also affect adhesion and their incorporation by amoebas.

Notwithstanding, other GV surface fibrils present some characteristics that distinguish them structurally from those of mimiviruses, such as *Cotonvirus japonicus*,[28] which has denser and shorter fibrils in its surface when observed by TEM analysis, which revealed a smoother arrangement when compared, for example, to lineage A mimiviruses' particles. As aforementioned, artifacts caused by preparation procedures cannot be excluded, but these structural disposition and size variations are also observed for orpheoviruses, which have larger particles than other mimiviruses (e.g. APMV) but present shorter fibrils.[29] In PBCV-1, its external fibrils extend from some of its capsomers and potentially facilitate particle attachment to algae hosts,[48] similar to the mimiviruses' mechanism initiating viral infection by attaching to these hosts' cell walls.

Current knowledge about fibrils: issues and needed studies

The absence of a three-dimensional (3D) solved fibril structure is still a limiting factor for further structural studies, including its binding mechanism onto host membranes during viral entry. Expression, purification, and obtaining crystallized or cryo-EM structures would help in predicting and performing subsequent binding or enzymatic assays. For instance, near-atomic resolution structures of different nucleocytoviruses,[49,50] yet one of the largest groups of viruses that infect eukaryotic hosts, could also help unravel possibilities for their proteins and assembly,[50] such as fibrils structural organization.

Moreover, in-depth studies about genes that have already been related to mimivirus fibrils[36] are also needed, as is the expansion in search of new genes, both in mimivirus and in other GV with fibrils, which remains a field to be explored. In addition, much of the existing information about these structures was obtained using mimiviruses as a study model, and it would be interesting if new studies with other giant viruses were carried out with the purpose of raising more data and knowledge about the possible diversity of fibrils. In this sense, it was shown that Megavirus chilensis has a gene cluster responsible for producing glycosylated proteins that can be associated with fibrils, and that this profile is shared with several other megaviruses,[51] still in need of more studies. Furthermore, if we consider L725, L829, and R135 as fibril-associated proteins,[32,36,37,39] the absence of them in some GV that presents fibrils (Table 1) could mean that other genes are responsible for the formation of fibrils in other species of nucleocytoviruses, and it is important to consider that perhaps this characteristic is an adaptive convergence in the evolution of GV.

AUTHORS' CONTRIBUTIONS

ILMdA, MGB, TBM, MSMS, and JSA wrote the review.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant number 303680/2022-9); Coordenação de Aperfeiçoamento

de Pessoal de Nível Superior/Ministério da Saúde (grant numbers 88882.348380/2010-1, 88887.595578/2020-00, and 88887.684031/2022-00); Fundação de Amparo à Pesquisa do Estado de Minas Gerais (PPM-00732-18); Pró-Reitoria de Pesquisa e de Pós-Graduação from Universidade Federal de Minas Gerais (04/2022); and Centro de Microscopia from Universidade Federal de Minas Gerais (1099). JSA is a Conselho Nacional de Desenvolvimento Científico e Tecnológico researcher (303680/2022-9).

ORCID ID

Jônatas Santos Abrahão https://orcid.org/0000-0001-9420-1791

REFERENCES

1. Koonin EV, Yutin N. Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology* 2010;**53**:284–92
2. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M. Distant mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci* 2011;**108**:17486–91
3. Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, Dornas FP, Trindade GS, Ferreira PP, Marques JT, Robert C, Raoult D, Kroon EG, La Scola B, Abrahão JS. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. *Virol J* 2014;**11**:95
4. Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic J-M, Ramus C, Bruley C, Labadie K, Shmakova L, Rivkina E, Couté Y, Abergel C, Claverie J-M. In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting Acanthamoeba. *PNAS* 2015;**112**:E5327–35
5. Bajrai LH, Benamar S, Azhar EI, Robert C, Levasseur A, Raoult D, La Scola B. Kaumoebavirus, a new virus that clusters with faustoviruses and Asfarviridae. *Viruses* 2016;**8**:278
6. Andrade ACDSP, Arantes TS, Rodrigues RAL, Machado TB, Dornas FP, Landell MF, Furst C, Borges LGA, Dutra LAL, Almeida G, Trindade GDS, Bergier I, Abrahão W, Borges IA, Cortines JR, de Oliveira DB, Kroon EG, Abrahão JS. Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. *Virol J* 2018;**15**:22
7. Boratto PVM, Serafim MSM, Witt ASA, Crispim APC, Azevedo BL, Souza GAP, Aquino ILM, Machado TB, Queiroz VF, Rodrigues RAL, Bergier I, Cortines JR, Farias ST, Santos RND, Campos FS, Franco AC, Abrahão JS. A brief history of giant viruses' studies in Brazilian biomes. *Viruses* 2022;**14**:191
8. Scheid P, Balczun C, Schaub GA. Some secrets are revealed: parasitic keratitis amoebae as vectors of the scarcely described pandoraviruses to humans. *Parasitol Res* 2014;**113**:3759–64
9. Zhang X, Xiang Y, Dunigan DD, Klose T, Chipman PR, Van Etten JL, Rossmann MG. Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid. *Proc Natl Acad Sci* 2011;**108**:14837–42
10. Yamada T, Onimatsu H, Van Etten JL. Chlorella viruses. *Adv Virus Res* 2006;**66**:293–336
11. Schulz F, Abergel C, Woyke T. Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat Rev Microbiol* 2022;**20**:721–36
12. Scola BL, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. A giant virus in amoebae. *Science* 2003;**299**:2033
13. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc Natl Acad Sci* 2009;**106**:21848–53
14. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 2013;**341**:281–6
15. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie J-M. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci* 2014;**111**:4274–9
16. Andreani J, Aherfi S, Bou Khalil JY, Di Pinto F, Bitam I, Raoult D, Colson P, La Scola B. Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. *Viruses* 2016;**8**:300
17. Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, Rossmann M, Colson P, Raoult D, La Scola B. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J Virol* 2015;**89**:6585–94
18. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, Sun S, Minsky A. Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. *PLoS Biol* 2008;**6**:e114
19. Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat Commun* 2018;**9**:749
20. Colson P, Ominami Y, Hisada A, La Scola B, Raoult D. Giant mimiviruses escape many canonical criteria of the virus definition. *Clin Microbiol Infect* 2019;**25**:147–54
21. Abrahão JS, Dornas FP, Silva LC, Almeida GM, Boratto PV, Colson P, La Scola B, Kroon EG. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. *Virol J* 2014;**11**:120
22. Colson P, La Scola B, Levasseur A, Caetano-Anollés G, Raoult D. Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat Rev Microbiol* 2017;**15**:243–54
23. Klose T, Herbst DA, Zhu H, Max JP, Kenttämaa HI, Rossmann MG. A mimivirus enzyme that participates in viral entry. *Structure* 2015;**23**:1058–65
24. Rodrigues RA, dos Santos Silva LK, Dornas FP, de Oliveira DB, Magalhães TF, Santos DA, Costa AO, de Macêdo Farias L, Magalhães PP, Bonjardim CA, Kroon EG, La Scola B, Cortines JR, Abrahão JS. Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. *J Virol* 2015;**89**:11812–9
25. Budin G, Chung HJ, Lee H, Weissleder R. A magnetic gram stain for bacterial detection. *Angew Chem Int Ed* 2012;**51**:7752–5
26. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG. Structural studies of the giant mimivirus. *PLOS Biol* 2009;**7**:e1000092
27. Andrade ACDSP, Rodrigues RAL, Oliveira GP, Andrade KR, Bonjardim CA, La Scola B, Kroon EG, Abrahão JS. Filling knowledge gaps for mimivirus entry, uncoating, and morphogenesis. *J Virol* 2017;**91**:e01335–10117
28. Takahashi H, Fukaya S, Song C, Murata K, Takemura M. Morphological and taxonomic properties of the newly isolated Cotonvirus japonicus, a new lineage of the subfamily Megavirinae. *J Virol* 2021;**95**:e0091921
29. Andreani J, Khalil JYB, Baptiste E, Hasni I, Michelle C, Raoult D, Levasseur A, La Scola B. Orpheovirus IHUMI-LCC2: a new virus among the giant viruses. *Front Microbiol* 2018;**8**:2643
30. Silva LK, dos S, Andrade AC, dos SP, Dornas FP, Rodrigues RAL, Arantes T, Kroon EG, Bonjardim CA, Abrahão JS. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. *Sci Rep* 2018;**8**:4000
31. Bajrai LH, Mougari S, Andreani J, Baptiste E, Delerce J, Raoult D, Azhar EI, La Scola B, Levasseur A. Isolation of Yasminevirus, the first member of Klosneuvirinae isolated in coculture with Vermamoeba vermiformis, demonstrates an extended arsenal of translational apparatus components. *J Virol* 2019;**94**:e01534–10119
32. Notaro A, Couté Y, Belmudes L, Laugieri ME, Salis A, Damonte G, Molinaro A, Tonetti MG, Abergel C, De Castro C. Expanding the occurrence of polysaccharides to the viral world: the case of mimivirus. *Angew Chem Int Ed* 2021;**60**:19897–904
33. Notaro A, Poirot O, Garcin ED, Nin S, Molinaro A, Tonetti M, De Castro C, Abergel C. Giant viruses of the Megavirinae subfamily possess

biosynthetic pathways to produce rare bacterial-like sugars in a clade-specific manner. *Microlife* 2022;**3**:uqac002
34. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M. The 1.2-megabase genome sequence of mimivirus. *Science* 2004;**306**:1344–50
35. Klose T, Kuznetsov YG, Xiao C, Sun S, McPherson A, Rossmann MG. The three-dimensional structure of mimivirus. *Intervirology* 2010;**53**:268–73
36. Boyer M, Azza S, Barrassi L, Klose T, Campocasso A, Pagnier I, Fournous G, Borg A, Robert C, Zhang X, Desnues C, Henrissat B, Rossmann MG, La Scola B, Raoult D. Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc Natl Acad Sci* 2011; **108**:10296–301
37. Villalta A, Schmitt A, Estrozi LF, Quemin ERJ, Alempic J-M, Lartigue A, Pražák V, Belmudes L, Vasishtan D, Colmant AMG, Honoré FA, Couté Y, Grünewald K, Abergel C. The giant mimivirus 1.2 Mb genome is elegantly organized into a 30-nm diameter helical protein shield. *Elife* 2022;**11**:e77607
38. Fischer D, Eisenberg D. Finding families for genomic ORFans. *Bioinformatics* 1999;**15**:759–62
39. Sobhy H, Gotthard G, Chabrière E, Raoult D, Colson P. Recombinant expression of mimivirus L725 ORFan gene product. *Acta Virol* 2017;**61**:123–6
40. de Aquino ILM, Serafim MSM, Machado TB, Azevedo BL, Cunha DES, Ullmann LS, Araújo JP, Abrahão JS. Diversity of surface fibril patterns in mimivirus isolates. *J Virol* 2023;**97**:e0182422
41. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 2004;**32**:W20–5
42. Siddiqui R, Khan NA. Biology and pathogenesis of Acanthamoeba. *Parasit Vectors* 2012;**5**:6
43. Dearborn DG, Korn ED. Lipophosphonoglycan of the plasma membrane of Acanthamoeba castellanii: FATTY ACID COMPOSITION. *J Biol Chem* 1974;**249**:3342–6
44. Rodrigues RAL, Abrahão JS, Drumond BP, Kroon EG. Giants among larges: how gigantism impacts giant virus entry into amoebae. *Curr Opin Microbiol* 2016;**31**:88–93
45. Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, Fournier PE. Mimivirus and the emerging concept of "giant" virus. *Virus Res* 2006;**117**:133–44
46. Van Etten JL, Lane LC, Meints RH. Viruses and viruslike particles of eukaryotic algae. *Microbiol Rev* 1991;**55**:586–620
47. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denef VJ, McMahon KD, Konstantinidis KT, Eloe-Fadrosh EA, Kyrpides NC, Woyke T. Giant virus diversity and host interactions through global metagenomics. *Nature* 2020;**578**:432–6
48. Agarkova IV, Lane LC, Dunigan DD, Quispe CF, Duncan GA, Milrot E, Minsky A, Esmael A, Ghosh JS, Van Etten JL. Identification of a chlorovirus PBCV-1 protein involved in degrading the host cell wall during virus infection. *Viruses* 2021;**13**:782
49. Fang Q, Zhu D, Agarkova I, Adhikari J, Klose T, Liu Y, Chen Z, Sun Y, Gross ML, Van Etten JL, Zhang X, Rossmann MG. Near-atomic structure of a giant virus. *Nat Commun* 2019;**10**:388
50. Shao Q, Agarkova IV, Noel EA, Dunigan DD, Liu Y, Wang A, Guo M, Xie L, Zhao X, Rossmann MG, Van Etten JL, Klose T, Fang Q. Near-atomic, non-icosahedrally averaged structure of giant virus Paramecium bursaria chlorella virus 1. *Nat Commun* 2022;**13**:6476
51. Piacente F, De Castro C, Jeudy S, Molinaro A, Salis A, Damonte G, Bernardi C, Abergel C, Tonetti MG. Giant virus Megavirus chilensis encodes the biosynthetic pathway for uncommon acetamido sugars. *J Biol Chem* 2014;**289**:24428–39

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Virology | Full-Length Text

The genomic and phylogenetic analysis of *Marseillevirus cajuinensis* raises questions about the evolution of Marseilleviridae lineages and their taxonomical organization

Bruna Luiza de Azevedo,[1] Victória Fulgêncio Queiroz,[1] Isabella Luiza Martins de Aquino,[1] Talita Bastos Machado,[1] Felipe Lopes de Assis,[1] Erik Reis,[1] João Pessoa Araújo Júnior,[2] Leila Sabrina Ullmann,[2] Philippe Colson,[3,4,5] Gilbert Greub,[6] Frank Aylward,[7,8] Rodrigo Araújo Lima Rodrigues,[1] Jônatas Santos Abrahão[1]

AUTHOR AFFILIATIONS See affiliation list on p. 17.

ABSTRACT Marseilleviruses (MsV) are a group of viruses that compose the Marseilleviridae family within the Nucleocytoviricota phylum. They have been found in different samples, mainly in freshwater. MsV are classically organized into five phylogenetic lineages (A/B/C/D/E), but the current taxonomy does not fully represent all the diversity of the MsV lineages. Here, we describe a novel strain isolated from a Brazilian saltwater sample named Marseillevirus cajuinensis. Based on genomics and phylogenetic analyses, M. cajuinensis exhibits a 380,653-bp genome that encodes 515 open reading frames. Additionally, M. cajuinensis encodes a transfer RNA, a feature that is rarely described for Marseilleviridae. Phylogeny suggests that *M. cajuinensis* forms a divergent branch within the MsV lineage A. Furthermore, our analysis suggests that the common ancestor for the five classical lineages of MsV diversified into three major groups. The organization of MsV into three main groups is reinforced by a comprehensive analysis of clusters of orthologous groups, sequence identities, and evolutionary distances considering several MsV isolates. Taken together, our results highlight the importance of discovering new viruses to expand the knowledge about known viruses that belong to the same lineages or families. This work proposes a new perspective on the *Marseilleviridae* lineages organization that could be helpful to a future update in the taxonomy of the Marseilleviridae family.

IMPORTANCE Marseilleviridae is a family of viruses whose members were mostly isolated from freshwater samples. In this work, we describe the first *Marseillevirus* isolated from saltwater samples, which we called *Marseillevirus cajuinensis*. Most of *M. cajuinensis* genomic features are comparable to other Marseilleviridae members, such as its high number of unknown proteins. On the other hand, *M. cajuinensis* encodes a transfer RNA, which is a gene category involved in protein translation that is rarely described in this viral family. Additionally, our phylogenetic analyses suggested the existence of, at least, three major Marseilleviridae groups. These observations provide a new perspective on Marseilleviridae lineages organization, which will be valuable in future updates to the taxonomy of the family since the current official classification does not capture all the Marseilleviridae known diversity.

KEYWORDS Marseillevirus, Marseilleviridae, lineages, phylogeny, taxonomy, evolution

Some years after the mimivirus discovery (1), the Marseilleviridae family became the second described group of large/giant viruses that are able to infect amoebas. The first Marseilleviridae isolate was obtained from water samples collected in a cooling tower in Paris, France (2). Typically, the members of the Marseilleviridae family exhibit

Editor Kristin N. Parent, Michigan State University, East Lansing, Michigan, USA

Address correspondence to Jônatas Santos Abrahão, jonatas.abrahao@gmail.com.

The authors declare no conflict of interest.

Received 19 March 2024
Accepted 19 April 2024
Published 16 May 2024

icosahedral particles with an average diameter of approximately 250 nm. The discovery of the first Marseillevirus (MsV) paved the way to describe several new isolates. Amoebae are likely marseillevirus hosts in the environment, and *Acanthamoeba* was the species used at the laboratory to isolate these viruses. Most of them were obtained from freshwater samples, including Lausannevirus, Tunisvirus, Melbournevirus, Cannes 8 virus, Noumeavirus, Brazilian marseillevirus, and Tokyovirus (3–9). Samples from other environments have been tested leading to the discovery of new Marseilleviridae isolates. Golden marseillevirus and Insectomime virus, for example, were isolated from golden mussels and internal organs of insect larvae, respectively (10, 11). In addition to several isolates already described, metagenomic studies have reported five divergent Marseilleviridae sequences detected in sediments from a hydrothermal vent (Loki's Castle), in the Atlantic Ocean (12). These discoveries in different types of environments, such as deep ocean regions, have expanded our understanding about the ecology of these viruses and locations where they can be isolated.

Marseilleviridae genomes are composed by circular double-stranded DNA molecules that range in size from 348 to 404 kbp (13). The G-C content ranges from 42.9% to 44.8%, and the number of predicted genes varies from 386 to 491 (13). Most part of Marseillevirus genomes encodes uncharacterized proteins. This is a general characteristic for giant viruses. Besides, Marseilleviruses do not have a diversity of genes involved in protein translation as observed in other giant viruses, such as those of Mimiviridae family (14). Although genes encoding translation factors have been described, it is not common to find in Marseilleviruses genome genes that encode transfer RNAs (tRNA) and aminoacyl-tRNA synthetases (13, 15). Considering tRNAs, they are currently described only in Tokyovirus and in the Loki's Castle metagenomic sequences (8, 12). As other Nucleocytoviricota phylum viruses, Marseillevirus genomes present a high mosaicism, which means that they have sets of genes of multiple origins (2).

The International Committee on Taxonomy of Viruses (ICTV) classifies all members of the Marseilleviridae family within the Nucleocytoviricota phylum, Pimascorivales order (16, 17). Currently, Marseilleviridae is composed by one genus (*Marseillevirus*), in which two species are included: *Marseillevirus marseillevirus* and *Senegalvirus marseillevirus*. There are two other species in the family, which are not included in any genus, the species *Lausannevirus* and *Tunisvirus* (16). However, a new organization and nomenclature for *Marseillevirus*-related taxa was recently proposed and is under approval for official publishing by ICTV. Phylogenetic analyses based on DNA polymerase proteins revealed the existence of five different Marseillevirus phylogenetic lineages, named A, B, C, D, and E. Lineage A is represented by the first isolate, *Marseillevirus marseillevirus* (2). Lineage B is represented by Lausannevirus, the second Marseilleviridae isolated (3). Lineages A and B have the largest number of isolates. The lineage C is represented by Tunisvirus and Insectomime virus (5, 17), while lineages D and E are composed of the Brazilian isolates, Brazilian marseillevirus and Golden marseillevirus, respectively (9, 11). Taking that into account, the known diversity of Marseilleviridae members is not being completely represented by the official taxonomy of the group. For example, the representatives of lineages D and E are not considered yet.

Almost 15 years after the discovery of the first *Marseillevirus*, it becomes apparent that the isolation of new viruses is important to elucidate the evolutionary history of this family. In this study, we report the discovery of *Marseillevirus cajuinensis*, a new isolate obtained from a saltwater sample from the Northeast coast of Brazil. This discovery paved the way for a genomic and phylogenetic characterization that suggested the existence of at least three major consistent Marseilleviridae groups. Analysis involving Clusters of Orthologous Groups (COGs), relative evolutionary distance (RED), and average amino acid/nucleotide identity (AAI and ANI) reinforces this three-group organization and helps to establish parameters for a future taxonomic organization of Marseilleviridae into three genera and different species. These findings provide a new perspective on Marseilleviridae lineages, which will be valuable in evolutive studies and future updates to the taxonomy of the family.

RESULTS

Marseillevirus cajuinensis: a new amoebae-infecting virus isolated from saltwater

From a saltwater sample collected in the Northeast coast of Brazil, we identified a new viral isolate able to infect *Acanthamoeba castellanii*. The transmission electronic microscopy (TEM) images containing infected amoebae showed icosahedral particles presenting shape similar to members of Marseilleviridae (Fig. 1A). With TEM analysis, it was also possible to observe some general morphological aspects from the isolate replication cycle. For example, we observed viral factories (VF) with an electron lucent aspect that occupied a large part of the cell cytoplasm (Fig. 1B, highlighted in pink). The images showed several viral particles in different maturation stages inside the VF (Fig. 1C, black arrow). Amorphic structures were also observed inside VFs (Fig. 1B through D), some of them horseshoe shaped (Fig. 1C and D, yellow arrows), resembling the crescent precursors of viral particles found in poxvirus and mimivirus viral factories (18, 19). Also, we observed giant vesicles harboring several viral particles at the end of the replication cycle. One of these giant vesicles reached a size of 5 µm (Fig. 1E, red arrow).

Besides TEM, the viral sample was analyzed by genome sequencing for a more accurate identification of the virus. Next-generation sequencing generated 225,574 reads that were assembled into a 166× depth single scaffold containing 380,653 bp. When compared with the National Center for Biotechnology Information (NCBI) database, the genome sequence best matched with a member of the Marseilleviridae family, presenting 79% of average nucleotide identity with Tokyovirus. Since we could confirm that the isolate is a Marseilleviridae, we named it *Marseillevirus cajuinensis*.

Marseillevirus cajuinensis genome

The circular double-stranded DNA molecule that composes *M. cajuinensis* genome has a G-C content of 45.24%. A total of 515 open reading frames (ORFs) were predicted that encode proteins with sizes ranging from 50 to 1,520 amino acids. GenBank sequence database searches suggested that 40 of the 515 proteins encoded by *M. cajuinensis* genome have functions related to DNA replication, recombination, and repair (Fig. 2A). This category includes a chaperone, a DNA topoisomerase, different nucleases, helicases, histones, and the DNA polymerase protein, which is commonly used as a marker for phylogeny. Other remarkable categories include signal transduction regulation and miscellaneous (Fig. 2A). The former primarily comprises serine/threonine protein kinases, while the latter consists of proteins whose function cannot be reliably predicted due to the presence of non-specific domains and/or repeats, such as ankyrin repeat-containing proteins and zinc finger proteins. *M. cajuinensis* genome also encodes for proteins involved in different metabolic processes, such as lipases and proteases. Furthermore, the major capsid protein (MCP) and the A32-like packaging ATPase are important proteins found in the virion structure and morphogenesis category.

More than half (59%) of the 515 *M. cajuinensis* proteins were classified as uncharacterized (Fig. 2A). Another 38 proteins were considered as ORFans since they had no hits with any other sequence in the used databases (Fig. 2A). Since several Marseilleviridae isolates have been described, it could be considered a high number of ORFans for a new isolate. However, by analyzing the ORFans' sizes, we observed that most of them (27/38) correspond to short polypeptides ranging from 50 to 100 amino acids (Fig. 2B). The possible high number of ORFans and their short sizes was intriguing. To confirm these results, a new gene prediction was performed using a different program. Thus, Prodigal was used instead of GeneMarkS. After this new analysis, 493 proteins (bigger than 50 amino acids) were predicted by Prodigal, 22 less than the 515 that were predicted by GeneMarkS. Interestingly, the number of predicted ORFans decreased considerably. Instead of 38, the new prediction returned only 9 ORFans. When analyzing the size of these newly predicted ORFans, it was possible to note that the number of ORFans bigger than 150 amino acids was increased, while the number of ORFans that were composed

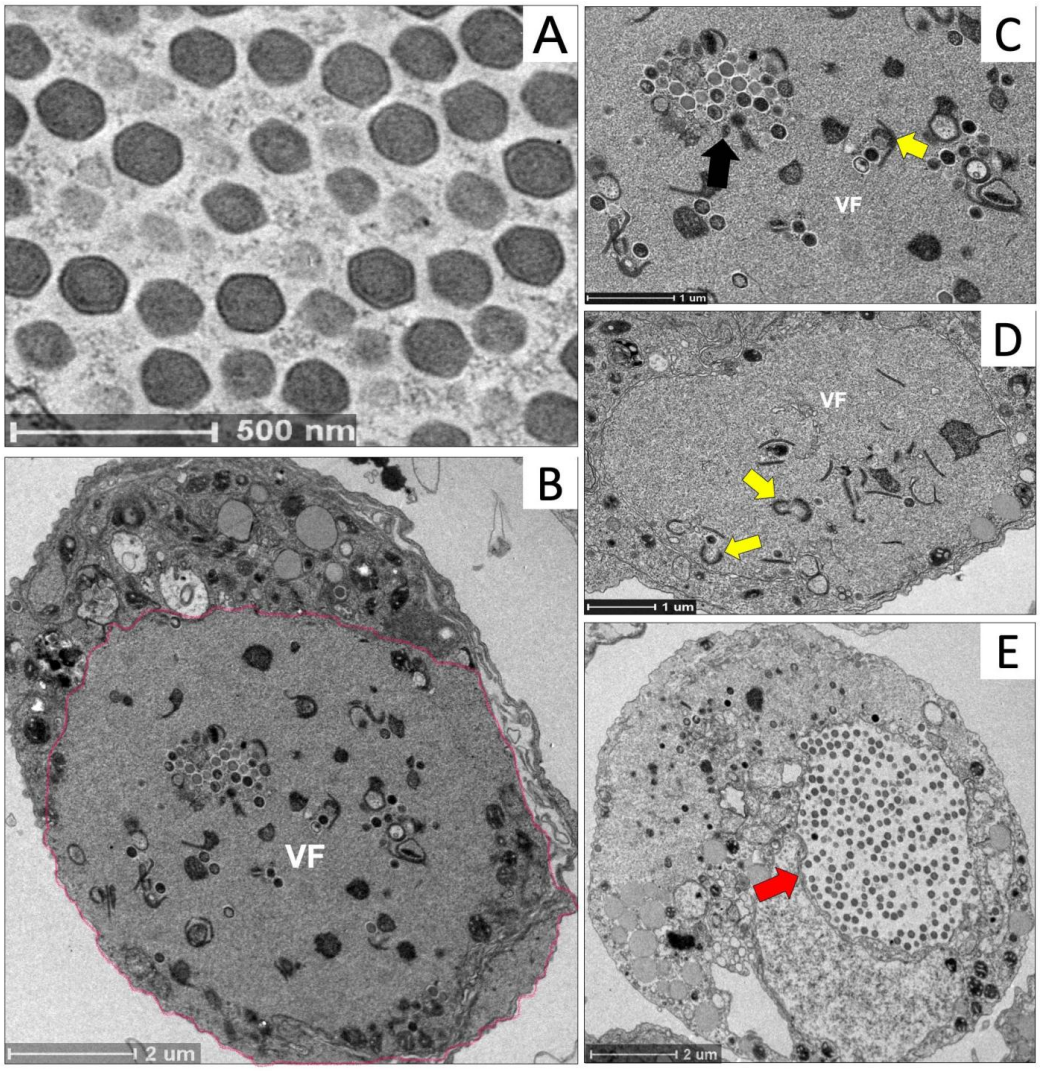

FIG 1 Transmission electronic microscopy images showing morphological characteristics from the new isolate inside *Acanthamoeba castellanii* cells. (A) Several icosahedral Marseillevirus-like particles. (B) *Acanthamoeba castellanii* cell containing a viral factory occupying a large portion of the cell. Viral factory boundaries are highlighted in pink. (C) A portion of Fig. 1B seen closer, showing a VF containing particles in different maturation stages (black arrow) and horseshoe-shaped structures (yellow arrow). (D) Another amoeba cell containing a VF filled with amorphic and horseshoe-shaped structures (yellow arrows). (E) Giant vesicle (red arrow) harboring several viral particles in a final stage of the cycle.

by less than 100 amino acids was considerably low (Fig. 2B). Here, predictions generated by GeneMarkS were selected to perform all the analysis in this work because this tool was the one the most consistently used for Marseilleviruses in previous works (2, 4–6, 9, 10, 17, 20).

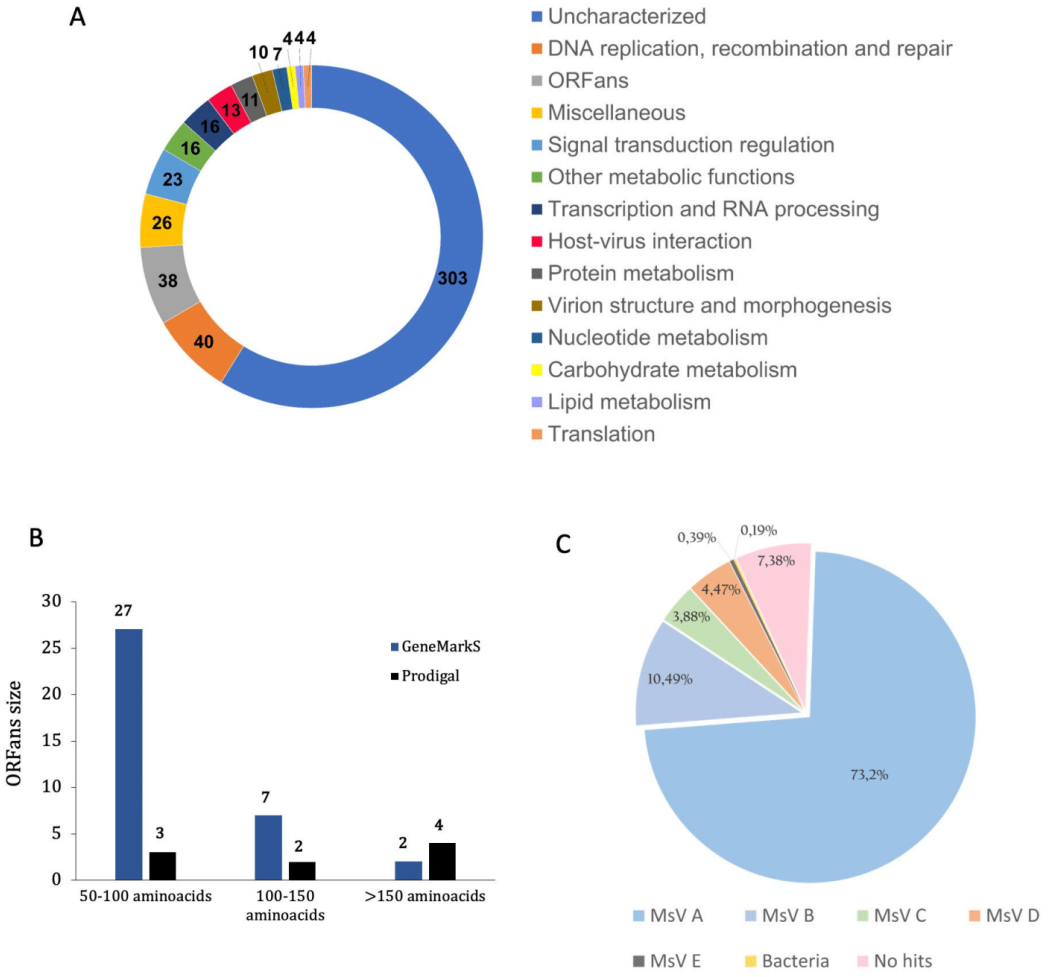

FIG 2 Genomic analysis of the *Marseillevirus cajuinensis* genome. (A) Number of *M. cajuinensis* proteins according to the function predicted during genome annotation. *n* = 515. (B) Comparison between the number of *M. cajuinensis* ORFans and their sizes when using different gene prediction tools. (C) Percentage of *Marseillevirus cajuinensis* proteins that presented best hit (BLASTp) with the five different Marseilleviridae lineages currently described (A–E). *n* = 515.

To know which organisms had their sequences considered as best hits with *M. cajuinensis* proteins, we analyzed hit by hit the results obtained in the BLASTp searches . Thus, for more than half (73.2%) of *M. cajuinensis* proteins, best hits from GenBank database, were Marseillevirus lineage A sequences (MsV A) (Fig. 2C). This suggests that *M. cajuinensis* could present a closer evolutionary relationship with lineage A, which can be confirmed by phylogeny. Also, this best hit analysis showed one *M. cajuinensis* protein (P86) whose best hit was a sequence encoded by a bacterium (Fig. 2C). P86 was the only protein for which no best hits were sequences from other members of the family Marseilleviridae.

Genome synteny analysis of different members of Marseilleviridae lineages highlighted the relatedness between *M. cajuinensis* and previously described lineage A members.

It is possible to observe that the organization of the *M. cajuinensis* genome blocks, separated and colored according to similarity, is more like that of the representative member of lineage A than of viruses from other lineages (Fig. 3). In this analysis, it is also possible to observe that in all the genomes analyzed, there is a region that appears to be more conserved, being approximately the last third of genome lengths (Fig. 3). Such conserved region in Marseilleviridae genomes was already described before and called a "core region" (21). More detailed synteny analysis containing different Marseilleviruses isolates from the five lineages can be found in Supplementary Figure 1 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-genomic-and-phylogenetic-analysis-of-marseillevirus-cajuinensis-raises-questions-about-the-evolution-of-marseilleviridae-lineages-and-their-taxonomical-organization.

Marseillevirus cajuinensis' translation-related genes and detection of tRNAs in different Marseilleviruses

The genomic analysis showed that *M. cajuinensis* encodes four different translation factors. No aminoacyl-tRNA-synthetase genes were found. Additionally, a search for transfer RNA sequences in *M. cajuinensis* genome was performed. To perform this search, two different programs (Aragorn and tRNAscanSE) were used. No tRNA sequence was found in *M. cajuinensis* genome by using tRNAscanSE. However, Aragorn was able to detect 1 tRNA sequence (tRNA-Gln-CTG) that has a 1,379-nucleotide intron.

Considering all the known giant/large amoeba-infecting viruses, tRNA encoding is not commonly described for Marseilleviridae isolates neither for other families from Pimascovirales. Otherwise, in groups phylogenetically related to the Pimascovirales order, such as cedratviruses and orpheoviruses, tRNAs were already described (22). Because of the difference in results between the two programs used to predict the tRNA, we carried out a search for tRNA in genomes from the five classical Marseilleviridae lineages that had complete sequences available in GenBank (March 2023). For this, we used both Aragorn and tRNAscanSE. The Aragorn program allows changing its parame-

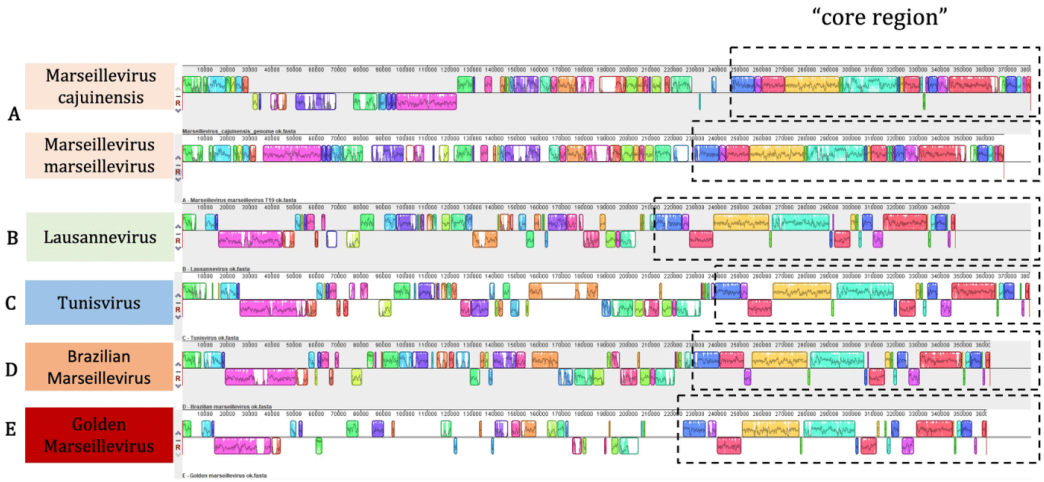

FIG 3 Genome synteny analysis of Marseilleviridae isolates representing the five currently described lineages and *Marseillevirus cajuinensis*. Each line represents the sequence of a different virus, which is identified in the legend on the left. The letters A, B, C, D, and E indicate the respective phylogenetic lineages of each analyzed virus. Blocks of the same color indicate similar regions between sequences. The areas without any colored blocks represent regions exclusive to that virus, that is, which do not show similarity with other viruses used in the analysis. Note: As they have a circular topology, the sequences were adjusted to start from the MCP aiming to facilitate interpretation of this figure. *Marseillevirus marseillevirus* was used as the reference genome.

ters to consider or not the presence of introns, and we performed the analysis in both conditions.

Thus, we detected tRNA sequences in Tokyovirus, *Marseillevirus marseillevirus*, Melbournevirus, Insectomime virus, Tunisvirus, and Golden marseillevirus using the Aragorn program with parameter allowing introns detection. When introns detection was not considered, it was only possible to detect tRNA in Tokyovirus, as the two sequences encoded by this virus do not have introns. Using tRNAscanSE, it was possible to detect tRNA only in the Tokyovirus sequence. Interestingly, tRNAscanSE detected three tRNA sequences in Tokyovirus, while Aragorn detected only two (see Supplementary Figure 2 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-genomic-and-phylogenetic-analysis-of-marseillevirus-cajuinensis-raises-questions-about-the-evolution-of-marseilleviridae-lineages-and-their-taxonomical-organization). Considering all tRNA sequences detected in different Marseillevirus, only two of the Tokyovirus tRNA were already described (8). Albeit little described in the Marseilleviruses of the five previously reported lineages, tRNAs have already been described in the sequences detected by metagenomics from samples from Loki's Castle. Although genomes assembled from metagenomes should be considered with caution, one of these sequences, called LCMAC202, was reported to encode 26 types of tRNA (12).

Phylogeny of different Nucleocytoviricota conserved proteins raises questions about Marseilleviridae lineages organization

To better elucidate the evolutionary relationship of *M. cajuinensis* with other Marseilleviridae members, phylogenetic analyses were performed using protein sequences that are considered conserved in Nucleocytoviricota. This set of conserved proteins includes the DNA polymerase, the A32-like packaging ATPase, and the late transcription factor VLT3 like, which were used as markers to construct phylogenetic trees (Fig. 4). It is noteworthy that the topology within lineages or genera varies not only among Marseilleviruses but also among other amoebal viruses, depending on the gene analyzed. Virus evolution is modular, with each gene subject to various nuances of a multitude of selective pressures.

Analyzing all phylogenies, it was possible to observe that *M. cajuinensis* groups together with sequences from Marseilleviruses of lineage A but represents a more divergent branch within this group (Fig. 4). Similar results were described for Tokyovirus within lineage A in previous works (8). It is noteworthy that *M. cajuinensis* and Tokyovirus cluster together in separate branch in the VLT3-like tree but not in all constructed trees, including the concatenated one. It is important to note that the divergence of *M. cajuinensis* within lineage A is comparable with the divergence that separates two different lineages (C and D). It raises questions about which criteria should be used to define what can be considered a new lineage or a new genus within the Marseilleviridae family. For example, by comparing lineage C and D branches, Brazilian marseillevirus is currently considered as a different lineage. Based on this, *M. cajuinensis*, and even Tokyovirus, could also be considered new lineages.

In addition to this questioning, it is also possible to observe the clear divergence of the common ancestor of the five classical Marseilleviridae lineages into three major groups: one that groups the current members of lineage A, corresponding to the current genus *Marseillevirus*, another that groups members of the current lineages B-C-D, and finally Golden marseillevirus (lineage E) in a third group. Noteworthy, the lineage E is closer from B-C-D branch than from the *Marseillevirus* genus (lineage A) but still presents a high divergence within its clade. Indeed, this same topology can be observed in a concatenated phylogenetic tree based on the three conserved sequences former analyzed individually (Fig. 4D).

Marseillevirus cajuinensis expands the pangenome of Marseilleviridae isolates

To understand the impact of the *M. cajuinensis* isolation on the pangenome and core genome of Marseilleviridae isolates, we searched for Clusters of Orthologous Groups of

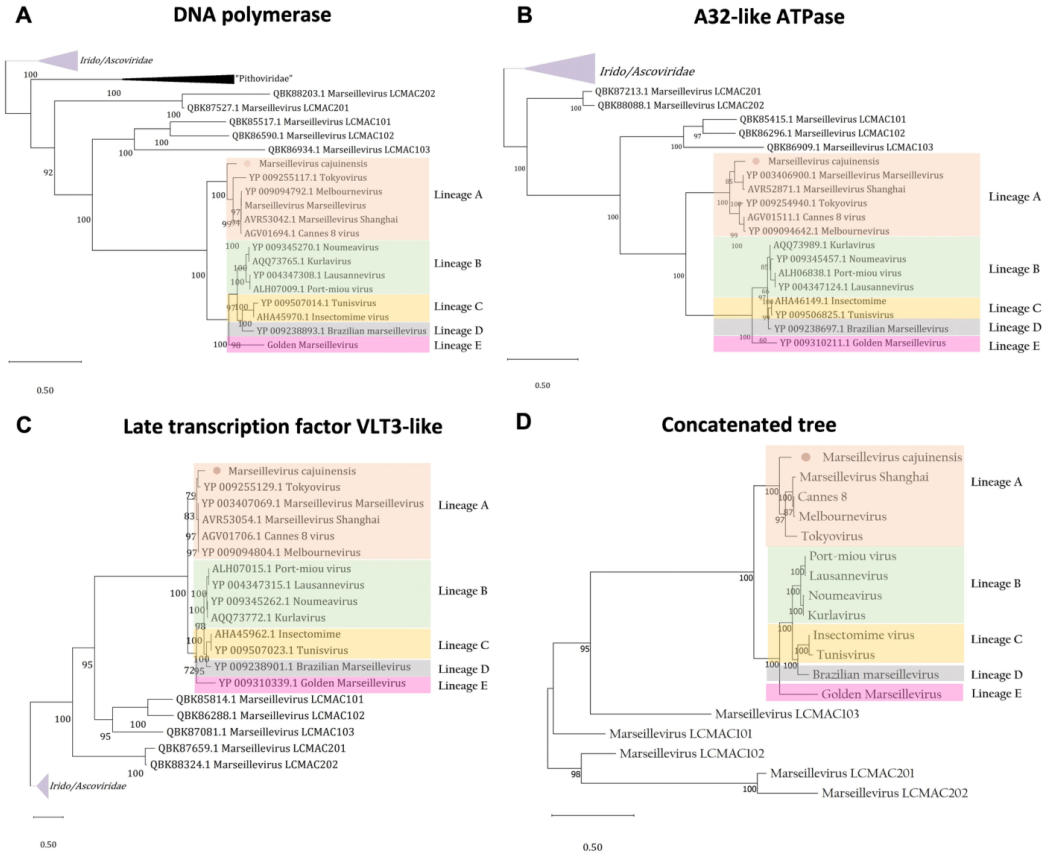

FIG 4 Marseilleviridae phylogeny using different conserved protein sequences. (A) Phylogeny based on DNA polymerase sequences. (B) Phylogenetic tree based on the A32-like packaging ATPase sequences. (C) Phylogeny based on the amino acid sequence of the VLT3-like late transcription factor sequences. (D) Concatenated sequence tree based on DNA polymerase, A32-like packaging ATPase, and VLT3-like late transcription factor sequences. *Marseillevirus cajuinensis* sequence is labeled in the tree with a pink disk. The trees were built using the maximum likelihood method, with statistical support based on 1,000 replicates (bootstrap). The best model, selected by IQtree (ModelFinder), for the trees was VT + F + I + G4 for (A), LG + G4 for (B), LG + I + G4 for (C), and VT + F + I + G4 for (D). The trees shown in A, B, and C were rooted on Iridoviridae branch as an outgroup. Concatenated tree was rooted at the midpoint. The tree scale bars represent the number of amino acid substitutions per site.

proteins shared between complete genome sequences of the isolates available in GenBank. Thus, it was possible to analyze the pangenome and core genome of isolated members of the Marseilleviridae family after including *M. cajuinensis* (Fig. 5) as well as the sharing of COGs between each lineage. It was observed that the pangenome of Marseilleviridae isolates increases from 598 to 1,626 when 13 new isolates are added in the analysis (Fig. 5). The first inserted sequence was Golden marseillevirus because it is the most divergent member of the five lineages. It was expected that with each new discovery of a different virus, the number of total COGs (pangenome) increases. This shows that the Marseilleviridae pangenome is still expanding and that the discovery of new additional viruses is warranted and important, as it will consequently lead to the discovery of new genes.

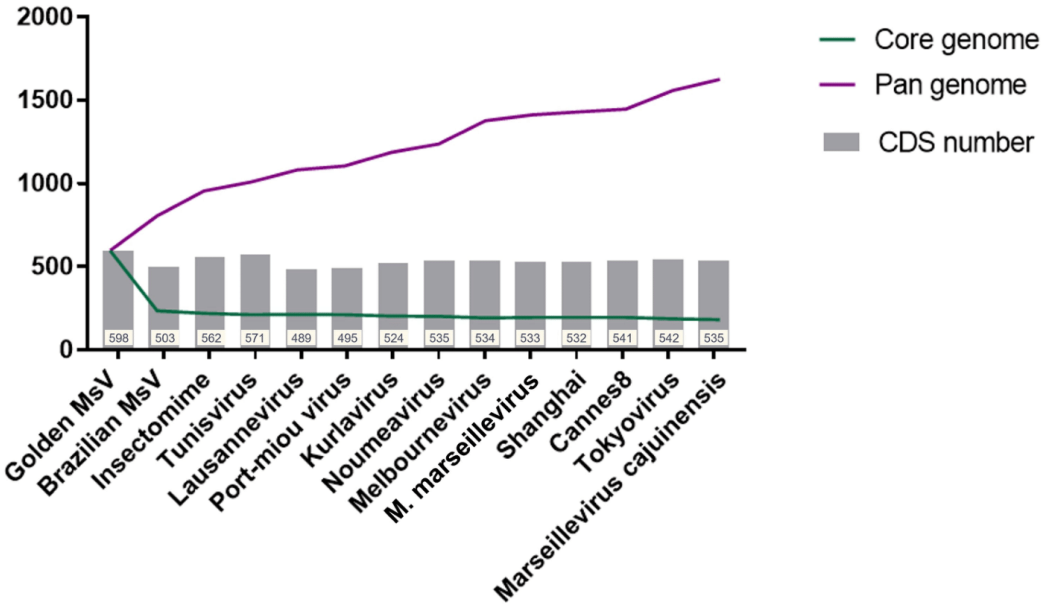

FIG 5 Analysis of the pangenome and core genome of Marseilleviridae isolates, including *Marseillevirus cajuinensis*. The curves indicate the variation in the number of Clusters of Orthologous Groups as new sequences are inserted in analysis. White boxes in the bottom of gray bars indicate the number of coding sequences (CDS) for each virus. Note: The number of CDS is based on the new gene prediction performed exclusively to this analysis.

On the other hand, the number of COGs shared by all the analyzed viruses (core genome) is 182 (Fig. 5). The graph analysis suggests that the core genome of Marseilleviridae isolates appears to have reached a plateau, suggesting that the gene content essential for the existence of these viruses is already relatively well defined, although the functions of most of these genes still need to be deciphered.

A detailed analysis of COGs shared between different Marseilleviruses reinforces the organization of Marseilleviridae in three major groups

In addition to pangenome and coregenome analyses, our data indicate the number of COGs that are shared between the members of the five Marseilleviridae classical lineages (Fig. 6). This analysis shows that *M. cajuinensis* (Fig. 6A, VI, red arrow) has 65 singletons, that is, clusters of proteins that are found only in its sequence. Among the members of lineage A (Fig. 6A, red disks), *M. cajuinensis* (VI, red arrow) and Tokyovirus (I) are those that have the greatest numbers of singletons. This reinforces the assumption that both viruses are the most divergent members of the lineage, corroborating the phylogeny results mentioned above. Altogether, the members of lineage A analyzed here share a total of 370 COGs that are unique to their lineage (Fig. 6A, red disks).

Among the other Marseilleviridae lineages analyzed, in lineage B (Fig. 6A, blue disks), it is possible to observe that all the four lineage members share 137 COGs that are absent in other lineages. Similarly, the lineage C members (Fig. 6A, purple disks) share 133 COGs that are exclusive to their lineage. On the other hand, the only known member of lineage D, Brazilian marseillevirus, has 52 singletons (Fig. 6A, gray arrow). This virus is described to compose its own lineage; however, in this analysis, it presents a smaller number of COGs than *M. cajuinensis* and Tokyovirus which are considered members of a same phylogenetic lineage (lineage A). Golden marseillevirus, the only known representative member of lineage E, has 294 singletons (Fig. 6A, orange arrow). The comparison of the

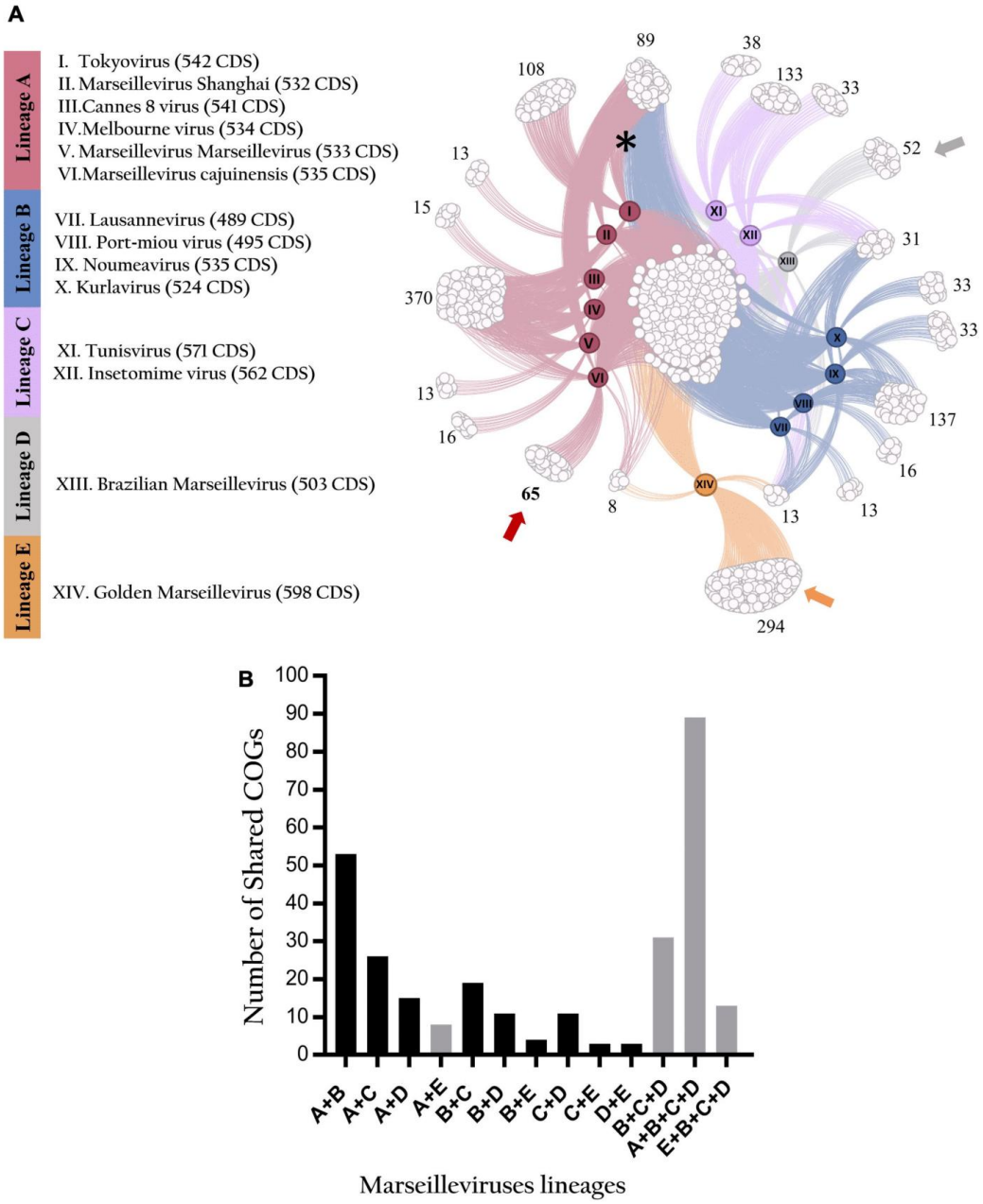

FIG 6 The sharing of Clusters of Orthologous Groups among different viruses belonging to the five classical lineages of Marseilleviruses. (A) Network showing the distribution of COGs among viruses. White circles represent COGs. The colored circles represent the analyzed viruses, separated by color, according to their respective phylogenetic lineages. Roman numerals individually identify each virus analyzed, while Arabic numbers indicate the number of COGs contained

(Continued on next page)

FIG 6 (Continued)

in each cluster. Note: The number of proteins obtained though GeneMarkS is indicated in parentheses for each virus. *Marseillevirus cajuinensis* is highlighted by a red arrow, Golden marseillevirus is highlighted by an orange arrow, Brazilian marseillevirus is highlighted by a gray arrow, and the sharing of COGs between lineages A, B, C, and D is highlighted by an asterisk. (B) Graph detailing the number of exclusive COGs shared between combinations of different classical Marseilleviridae lineages. The black bars correspond to combinations of lineages that are not represented in Fig. 6A, while the gray bars correspond to combinations already represented in Fig. 6A. Note: Some of the lineage's combinations were not represented in Fig. 6A because some of them often overlap themselves in the network, making analysis hard.

number of exclusive shared COGs between the different lineages is detailed in Fig. 6B. In this graph, it is possible to analyze the sharing of COGs between different lineages in a simpler way. Lineages A and E (A + E), for example, share only eight COGs. This number decreases when comparing the other lineages with lineage E. Lineages B and E (B + E) share only four COGs, while lineages C and E (C + E), and D and E (D + E) share only COGs. Together, lineages A, B, C, and D (A + B + C + D) share 89 COGs that are not found in lineage E (Fig. 6A, asterisk). Thus, these data reinforce the divergence of lineage E among Marseilleviridae and support its assignment in a distinct group of the family.

Using the data obtained in COGs analysis described above, a hierarchical clustering phenophyletic tree was constructed based on the presence and absence of COGs in the analyzed sequences (Fig. 7). In this figure, it was possible to observe a topology very similar to what was observed in the phylogenetic trees described in this work. Representatives of lineage A are organized into a group (group I) that represents the current genus *Marseillevirus* and apart from representatives of lineages B, C, D (group II), and E (group III) (Fig. 7).

Within lineage A, it was possible to observe the organization of the viruses into two subgroups, one was composed of *M. cajuinensis* and Tokyovirus, and the other was composed of the other lineage A viruses (Fig. 7). The same happens in group II since there are three subgroups composed by each member of lineages B, C, and D

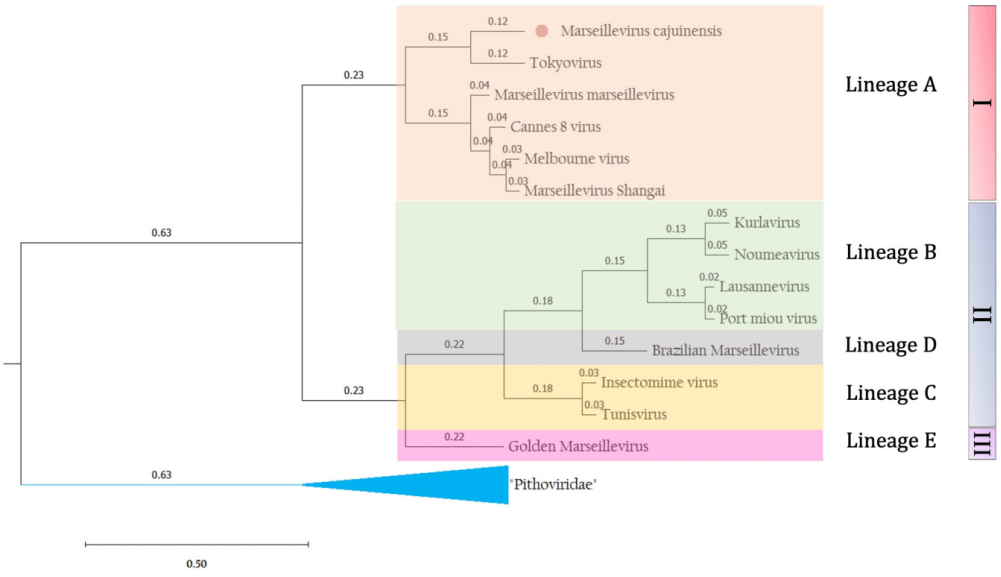

FIG 7 Hierarchical clustering tree considering the presence and absence of Clusters of Orthologous Groups in different viruses of Marseilleviridae family. *Marseillevirus cajuinensis* is labeled in the tree with a pink circle. Scale bar represents arbitrary values that express the evolutionary distance based on presence-absence of COGs.

(Fig. 7). These results reinforce what was observed in phylogeny and genomic analyses. Also, they reinforce the questions raised about the classical organization of Marseilleviridae viruses in five lineages. For example, if the tree branches (lineages B, C, and D) that compose group II are represented by viruses originally classified into three different lineages, this could justify that *Marseillevirus cajuinensis* and Tokyovirus could be classified in its own lineages. On the other hand, the different branches that compose the classical Marseilleviridae lineages could also be considered as members of major groups.

Relative evolutionary distance and average amino acid/nucleotide identity analyses suggest the organization of Marseilleviridae isolates in three putative genera

To quantify all these observations and make them more consistent, an RED analysis was performed. Thus, we considered three groups for this analysis based on the three major clades observed in DNA polymerase phylogeny: group I (lineage A), group II (lineages B, C, D), and group III (lineage E) (Fig. 8A). DNA polymerase phylogeny was selected because it is the main phylogenetic marker for Nucleocytoviricota, but the same topology was observed for all phylogenies (Fig. 4) and for the hierarchical clustering tree as well (Fig. 7).

RED values varies from 0 to 1, and threshold values for different taxonomic levels in Nucleocytoviricota were defined previously (23). The previously reported RED values for genus ranged from 0.69 to 0.995 (23). The present analysis showed that group I (lineage A) had a RED value of 0.86, while group II (lineages B, C, and D) had a RED value of 0.83 (Fig. 8A). These numbers are consistent with values expected for the genus level (23). Because group III (lineage E) is composed by a single genome (Golden marseillevirus), RED analysis cannot be performed. However, when Golden marseillevirus sequence is included in group II (lineages B, C, and D), the RED value decreases to 0.78. Although this value could still classify group II as a genus while including Golden marseillevirus, the lower RED value makes group II less consistent when comparing with group I. Thus, excluding lineage E from group II and considering it as a separate group were most consistent here.

Additionally, sequences of the Marseillevirus isolates previously analyzed in this work were submitted to an average nucleotide identity analysis. The ANI analysis delineated three main groups of Marseilleviruses, considering an ANI cutoff >75% (Fig. 8B). This value corroborates with our phylogenetic analyses and hierarchical clustering of COGs, revealing the existence of three distinct groups in Marseilleviridae, possibly corresponding to three distinct genera. Within groups, we can use pairwise ANI >95% to define viral species, as used for members of Imitervirales (24). In this case, we can define eight viral species, among which three belong to group I—one of them corresponding to the *Marseillevirus cajuinensis*; four belong to group II; and only one, consisting of the Golden marseillevirus isolate, belongs to group III. Also, an average amino acid identity analysis was performed. The same three groups could be observed in the bidirectional AAI analysis, with AAI >65% (Fig. 8B). Moreover, in the AAI estimation, we clearly saw eight putative viral species, with AAI >95%.

DISCUSSION

In this work, we describe the isolation of *M. cajuinensis* from a saltwater sample from the Northeast coast of Brazil. Although members of the family Marseilleviridae have already been isolated from samples of different sources (2, 10, 11, 17, 25), most of them were obtained from freshwater or mud from rivers and lakes (2–4, 6–8). To our knowledge, this is the first time that a Marseilleviridae member has been isolated from ocean water, although they have already been detected in this type of environment through metagenomic analyses (12).

TEM images revealed that the replication cycle of *M. cajuinensis* shares similar characteristics with other members of the Marseilleviridae family, such as the presence of

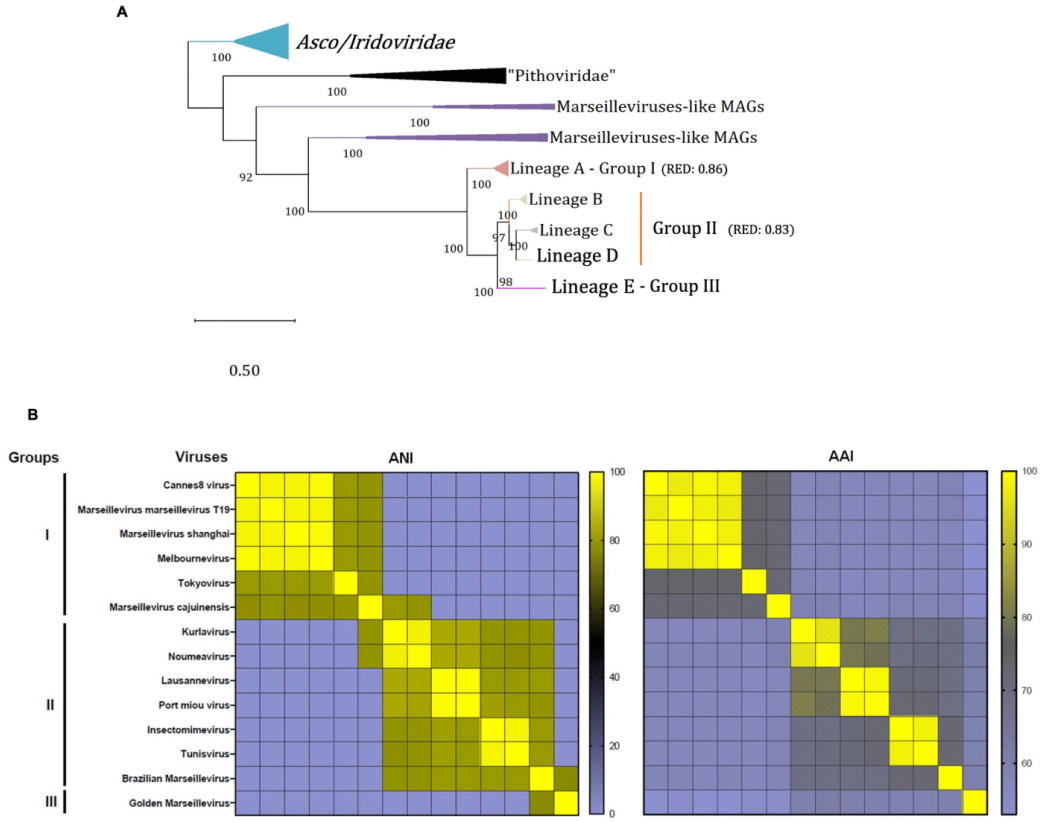

FIG 8 (A) DNA polymerase phylogenetic tree illustrating the three-group proposal for Marseilleviridae. RED values are indicated for groups I and II, and the numbers are consistent with the genus level for taxonomy. Group III does not have a RED value because there is only one virus in lineage E. The tree was built using the maximum likelihood method, with statistical support based on 1,000 replicates (bootstrap). The best model, selected by IQtree (ModelFinder), for the tree was VT + F + I + G4. The tree was rooted on Iridoviridae branch as an outgroup. The tree scale bars represent the number of amino acid substitutions per site. Note: MAGs, metagenome-assembled genomes. (B) Average nucleotide identity and average amino acid identity analysis of Marseilleviridae. Fourteen Marseilleviruses are grouped based on a similarity matrix composed by ANI (left heatmap) and AAI (right heatmap). The three viral groups are indicated. ANI <75% was set to zero. ANI values ranged from 77 to 100, and AAI values ranged from 53 to 100.

giant vesicles by the end of the cycle (26). These vesicles are important structures for their replication cycles, as the viruses can be released from the cell inside these vesicles, referred to as "expelled vesicles" (27, 28). This mechanism is important to ensure greater effectiveness in the subsequent entry of particles into another amoeba to start new cycle, corresponding to a sort of Trojan horse strategy. Alternatively, such structure may be related to the exocytosis of the viral progeny (26) and their increased resistance to harsh environments when being outside amoebae, waiting for new hosts. The presence of amorphic and horseshoe-like structures inside the viral factory (Fig. 1C and D) requires further investigations, and more detailed analyses of the replication cycle are needed to infer a biological function to these structures. It is known that in the initial stages of Marseilleviridae isolates viral factory formation, endosomal membranes are recruited. These recruited membranes are involved in the formation of the internal membranes that compose viral capsids (26). Therefore, it can be hypothesized that *M. cajuinensis* VF

structures have been recruited from some cellular component to perform a function that is not yet known.

After analyzing the genome of *M. cajuinensis*, we observed that its genome size, GC content, and number of genes are compatible with that described for other members of the Marseilleviridae family (2, 3, 5, 9, 11, 13). The functional categorization of proteins suggests that more than half of the *M. cajuinensis* genome encode proteins with unknown functions. This is a very common characteristic among giant amoeba viruses and reinforces the need for new studies that aim the elucidation of the unknown functions of these proteins (1, 29, 30). Also, we detected 38 ORFans in a first gene prediction, but most of them have sizes that range from 50 to 100 amino acids. This puts in doubt whether these sequences are real proteins or whether they are artifacts due to the main gene prediction protocol used in this study. When using a new gene prediction tool, the number of ORFans decreased to 9. This considerable difference in results might be due to the lack in updates of prediction tools that currently are mostly indicated for prokaryotes, eukaryotes, and for viruses that have smaller and fewer complex genomes. The number of ORFs was also affected when different parameters and tools for prediction are utilized. In addition, we detected a gene that encodes a tRNA, which is not commonly described in members of the family Marseilleviridae (13, 14). By analyzing sequences from other Marseilleviridae members using different programs and parameters, we detected tRNAs in five Marseilleviruses that, to our knowledge, have not been previously described as encoding such sequences. The absence of tRNA detection might be attributed to the use of a single algorithm in most previous analyses. Our current analysis shows a difference in the sensitivity of tRNA detection, with Aragorn being more sensitive than the tRNAscanSE algorithm tested in parallel. This is likely due to the different tRNA search models and parameters used by each tool (31, 32). For example, the Aragorn does not depend on the taxonomic lineage specification as parameter to achieve maximum search sensitivity, whereas the tRNAscan-SE does (31). Thus, current gene prediction algorithms lack updates that consider the singularities of giant viruses. This highlights the importance to systematically use more than one algorithm for ORFs and tRNA prediction, as they can complement each other and can stimulate deeper investigations.

Comparisons between *M. cajuinensis* and members of the *Marseillevirus* genus, conducted through genomic and phylogenic analyses, showed that they are all grouped within the lineage A. Despite this, *M. cajuinensis* and Tokyovirus form a divergent branch within this lineage. Phylogenetic analyses also showed that the common ancestor of the five classical Marseilleviridae lineages is in fact diversified into three main branches, which we refer to as group I (lineage A), group II (lineages B, C, and D), and group III (lineage E). After analyzing parameters such as RED, AAI, and ANI, it is possible to suggest that these three groups could potentially be considered as three genera. Taxonomically, group I currently corresponds to *Marseillevirus* genus, and group II corresponds to the recently proposed *Losanna* genus (that includes the *L. lausannense* and the *L. tunisiense* species). Group III and some members of the other proposed groups (e.g. Brazilian marseillevirus) remain to be officially assigned taxonomically. Also, AAI and ANI analyses suggested the organization of Marseilleviridae isolates in eight species (see Supplementary Figure 3 at https://www.giantviruses.com/sup-material-of-papers/sup-material-the-genomic-and-phylogenetic-analysis-of-marseillevirus-cajuinensis-raises-questions-about-the-evolution-of-marseilleviridae-lineages-and-their-taxonomical-organization), according to the percentage of sequence identity between each other. This could be helpful to classify the isolates that are still not considered officially.

The analyses of COGs shared between different lineages highlight conserved and variable COGs in each lineage. An in-depth analysis to understand the function of each protein belonging to the clusters was not performed here, and most part of these proteins might not have a known function. However, it is possible to hypothesize that conserved COGs might represent important genes to the lineage, possibly inherited

from their common ancestor. Conversely, the proteins belonging to clusters that vary among lineages might have different origins. The presence/absence of COGs analysis complemented the phylogeny and reinforced both the greater divergence of *Marseillevirus cajuinensis* within lineage A and the organization of Marseilleviridae classical lineages into three groups. However, is it worth mentioning that defining the number of Marseilleviridae lineages/genera is a big challenge and might be treacherous as it depends on the methods and the viruses considered in the analysis. Taking that into account, it is clear that there is a need to continue efforts to obtain new isolates as the Marseilleviridae pangenome is still open. The raised questions about the number of lineages within the Marseilleviridae family reflect the impact of new virus discovery on taxonomists' perspectives, as these new strains add new information to the analyses, sometimes leading to different tree topologies. Such new topologies call for updating phylogenetic organization and taxonomy to ensure that genus and species taxonomic levels better reflect the reality of diversity in a given taxon, rather than being biased due to sampling mostly specific habitats, such as freshwater. For this taxonomic update to happen, it is necessary to first establish the parameters that are needed to classify these viruses into new species or genus. Thus, this work represents a contribution to shape future updates in Marseilleviridae taxonomy.

MATERIALS AND METHODS

Viral isolation, multiplication, purification, and titration

The isolate was obtained through the collection of saltwater samples at Cajueiro da Praia city, located in Piaui state (Northeast coast of Brazil). The protocol was based on the inoculation of the collected samples on 96-well plates containing *Acanthamoeba castellanii* cells (33). The inoculated wells that presented cytopathic effects (i.e., rounding cells and cellular lysis) had their content collected and analyzed through transmission electron microscopy. After confirming the isolation, the virus was inoculated at a multiplicity of infection (MOI) of 0.01 in cell culture flasks containing 1.4×10^7 *Acanthamoeba castellanii* cells and 35 mL of peptone-yeast extract-glucose (PYG) medium, supplemented with penicillin (100 U/mL; Cellofarm, Brazil), streptomycin (100 µg/mL; Sigma-Aldrich, USA), and amphotericin B (0.25 µg/mL; Cultilab, Brazil). The cells were incubated at 32°C. Non-infected cells maintained in the same conditions were used as control. When viral-induced cytopathic effects were observed, the flask's content was collected. This content was filtered through 0.45 µm pores, and then it was ultracentrifuged (36,000 x g) in a 25% sucrose cushion for 2 hours. The pellet containing purified viral particles was homogenized in 300 µL of phosphate-buffered saline (PBS 1×). All the viral titers were obtained and calculated using the end-point method (34).

Transmission electron microscopy

To analyze the morphology of isolated viral particles, the samples were prepared for TEM. First, 7×10^6 *A. castellanii* cells, cultured in 25 mL of PYG medium, were inoculated with the virus at an MOI of 0.01. Once cytopathic effects were observed, we performed two consecutive washes with 0.1 M sodium phosphate buffer, and we subsequently fixed the cells for 2 hours at room temperature under rotation in an orbital mixer. The fixation solution consisted of 0.1 M sodium phosphate buffer and 2.5% glutaraldehyde. Following this initial fixation step, the cells underwent a secondary fixation with 2% osmium tetroxide before being embedded in Epon resin. This resin allowed an ultramicrotomy, and the 60-nm-thick sections were then examined using a transmission electron microscope (Spirit Biotwin FEI-120 kV) at the Center of Microscopy of the Federal University of Minas Gerais (CM-UFMG).

Sequencing, assembly, and annotation

The purified virus was sequenced using an Illumina MiSeq instrument with a paired-end library using the Illumina DNA Prep Kit (Illumina Inc., San Diego, CA, USA). The FastQC program was used for quality control of the obtained reads, and the per base sequence quality was considered satisfactory (phred >28). The reads were trimmed using the Trimmomatic tool (35). Genome *de novo* assembly was performed using Spades 3.12 program with default parameters (36, 37). The obtained scaffold was compared with sequences from the NCBI database, using BLASTn (database: nr/nt; expect threshold: 10^{-3}). Open reading frames were predicted with the GeneMarkS tool and Prodigal (38, 39), considering only proteins that were bigger than 50 amino acids. Additionally, tRNA-coding sequences (CDSs) were predicted using ARAGORN (parameters –type: tRNA; allow intron: yes and no (alternately); topology: circular; strand- both) and tRNAscanSE (parameters - Sequence source: general tRNA model; Search mode: default, Genetic Code for tRNA Isotype Prediction: universal) (31, 32). ORFs were annotated using BLASTp (expect threshold: 10^{-3}) against the NCBI non-redundant protein sequence (nr) database aiming to search for similar sequences in this database. The functional categorization of predicted proteins was carried out based on the Nucleo-Cytoplasmic Virus Orthologous Groups (40, 41).

Synteny and phylogenetic analysis

To perform synteny analyses, genome sequences of different MsV isolates were obtained from the NCBI GenBank database. Only genomes from isolated Marseilleviruses (excluding those built from metagenomic data) and that were complete and available in GenBank (March 2023) were selected. As they have a circular topology, the sequences were manually curated to start from the major capsid protein aiming to facilitate image interpretation. After curating the sequences, synteny analysis was performed using the MAUVE program, with default parameters (42). The following genome sequences were retrieved from GenBank and then analyzed: Tokyovirus (NC_030230.1); *Marseillevirus marseillevirus* (GU071086.1); Cannes 8 virus (KF261120.1); Marseillevirus Shanghai (MG827395.1); Melbournevirus (KM275475.1); Kurlavirus (KY073338.1); Lausannevirus (HQ113105.1); Noumeavirus (KX066233.1); Port-miou virus (KT428292.1); Insectomime (HG428764.1); Tunisvirus (KF483846.1); Brazilian MsV (KT752522.1); Golden MsV (KT835053.1).

Phylogenetic trees were constructed using the IQtree software (version 1.6.12) with 1,000 bootstrap replicates as branch support (43). To prepare the data sets for alignment, a search for similar sequences was performed using the NCBI non-redundant protein sequences (nr) database and BLASTp with an expected threshold of 10^{-3}. Sequence alignment was performed using the MUSCLE algorithm (44). The best-fit substitution models were determined using the ModelFinder algorithm within IQtree (45). Finally, the resulting phylogenetic trees were visualized and edited using MEGA X software (46).

Relative evolutionary distance analyses were performed using phylogeny constructed according to parameters mentioned above. RED values were calculated using the R package "castor" (47), and the thresholds for taxonomic levels were defined as described in the Results section based in a previous work (23).

Pangenome and COGs analysis

All complete MsV sequences that were previously obtained for synteny analyses from GenBank were also subjected to a new gene prediction using GeneMarkS (38). The amino acid sequences of each predicted CDS were analyzed using the ProteinOrtho software (parameters - selfblast, identity: 30%, coverage: 50%, and *e*-value of 10^{-5}) (48). The output files generated by ProteinOrtho were used to analyze the pangenome and core genome of isolated MsV. Also, output files of orthologous proteins were used to compare the number of Clusters of Orthologous Groups that are shared between the studied viruses and to construct a hierarchical clustering based on the presence and absence of COGs

in different MsV. To analyze the sharing of COGs between different MsV, a network representation was constructed using the Gephi 0.10 software. For this, data obtained in ProteinOrtho analysis were used to create spreadsheets containing the "nodes" (viruses and COGs) and the "edges" (presence of COGs in each virus). Network representation was built using an algorithm based on attraction and repulsion forces (Force Atlas). To perform COGs presence and absence analysis, a binary file was generated, and a phenetic tree was created in the MultiExperiment Viewer program, version 4.9.0, using the hierarchical clustering algorithm and the Pearson correlation as distance metric (49).

Average nucleotide and amino acid identities

Whole-genome average nucleotide identity analysis was performed using FastANI (50) implemented on Galaxy Server (https://usegalaxy.eu/), on the complete genomes of 14 Marseilleviruses obtained inform the NCBI GenBank database. ANI <75% was considered 0. Average amino acid identity was calculated using reciprocal best hits (two-way AAI) between two Marseilleviruses' protein genomic data sets, considering an identity cutoff of 20%. AAI was estimated using the AAI calculator (http://enve-omics.ce.gatech.edu/aai/).

ACKNOWLEDGMENTS

We would like to thank all colleagues from Grupo de Estudo e Prospecção de Vírus Gigantes (GEPVIG) and from Laboratório de Vírus of Universidade Federal de Minas Gerais (UFMG). Also, we thank Centro de Microscopia of UFMG for the contribution of microscopy images.

We thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), and FAPEMIG (Fundação de Amparo à Pesquisa do estado de Minas Gerais) for research funding.

This research is registered at SISGEN and SISBIO.

J.S.A., J.P.A.J., and R.A.L.R. are CNPq researchers.

P.C., G.G., and J.S.A. are contributors of the ICTV Marseillevirus study group (2016-2023).

AUTHOR AFFILIATIONS

[1]Laboratório de Vírus, Departamento de Microbiologia, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil
[2]Laboratório de Virologia, Departamento de Microbiologia e Imunologia, Instituto de Biotecnologia, Universidade Estadual Paulista (Unesp), Alameda das Tecomarias s/n, Chácara Capão Bonito, Botucatu, Brazil
[3]IHU Méditerranée Infection, Marseille, France
[4]Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille University, Marseille, France
[5]Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille, France
[6]Centre for Research on Intracellular Bacteria and Giant Viruses, Institute of Microbiology, University Hospital Centre and University of Lausanne, Lausanne, Switzerland
[7]Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA
[8]Center for Emerging, Zoonotic, and Arthropod-Borne Infectious Disease Virginia Tech, Blacksburg, Virginia, USA

AUTHOR ORCIDs

Gilbert Greub http://orcid.org/0000-0001-9529-3317
Frank Aylward http://orcid.org/0000-0002-1279-4050
Rodrigo Araújo Lima Rodrigues http://orcid.org/0000-0001-7148-4012
Jônatas Santos Abrahão http://orcid.org/0000-0001-9420-1791

DATA AVAILABILITY

The *M. cajuinensis* genome sequence is available in GenBank under accession number OR991738.

REFERENCES

1. Scola BL, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D. 2003. A giant virus in amoebae. Science 299:2033–2033. https://doi.org/10.1126/science.1081867
2. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D. 2009. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A 106:21848–21853. https://doi.org/10.1073/pnas.0911354106
3. Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, Croxatto A, Greub G. 2011. Lausannevirus, a giant amoebal virus encoding histone doublets. Environ Microbiol 13:1454–1466. https://doi.org/10.1111/j.1462-2920.2011.02446.x
4. Aherfi S, Pagnier I, Fournous G, Raoult D, La Scola B, Colson P. 2013. Complete genome sequence of Cannes 8 virus, a new member of the proposed family "Marseilleviridae". Virus Genes 47:550–555. https://doi.org/10.1007/s11262-013-0965-4
5. Aherfi S, Boughalmi M, Pagnier I, Fournous G, La Scola B, Raoult D, Colson P. 2014. Complete genome sequence of Tunisvirus, a new member of the proposed family *Marseilleviridae*. Arch Virol 159:2349–2358. https://doi.org/10.1007/s00705-014-2023-5
6. Doutre G, Philippe N, Abergel C, Claverie J-M. 2014. Genome analysis of the first *Marseilleviridae* representative from Australia indicates that most of its genes contribute to virus fitness. J Virol 88:14340–14349. https://doi.org/10.1128/JVI.02414-14
7. Fabre E, Jeudy S, Santini S, Legendre M, Trauchessec M, Couté Y, Claverie J-M, Abergel C. 2017. Noumeavirus replication relies on a transient remote control of the host nucleus. Nat Commun 8:15087. https://doi.org/10.1038/ncomms15087
8. Takemura M. 2016. Morphological and taxonomic properties of *Tokyovirus*, the first *Marseilleviridae* member isolated from Japan. Microbes Environ 31:442–448. https://doi.org/10.1264/jsme2.ME16107
9. Dornas FP, Assis FL, Aherfi S, Arantes T, Abrahão JS, Colson P, La Scola B. 2016. A Brazilian marseillevirus is the founding member of a lineage in family *Marseilleviridae*. Viruses 8:76. https://doi.org/10.3390/v8030076
10. Colson P, Fancello L, Gimenez G, Armougom F, Desnues C, Fournous G, Yoosuf N, Million M, La Scola B, Raoult D. 2013. Evidence of the megavirome in humans. J Clin Virol 57:191–200. https://doi.org/10.1016/j.jcv.2013.03.018
11. Dos Santos RN, Campos FS, Medeiros de Albuquerque NR, Finoketti F, Côrrea RA, Cano-Ortiz L, Assis FL, Arantes TS, Roehe PM, Franco AC. 2016. A new marseillevirus isolated in Southern Brazil from *Limnoperna fortunei*. Sci Rep 6:35237. https://doi.org/10.1038/srep35237
12. Bäckström D, Yutin N, Jørgensen SL, Dharamshi J, Homa F, Zaremba-Niedwiedzka K, Spang A, Wolf YI, Koonin EV, Ettema TJG. 2019. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. mBio 10:e02497-18. https://doi.org/10.1128/mBio.02497-18
13. Sahmi-Bounsiar D, Rolland C, Aherfi S, Boudjemaa H, Levasseur A, La Scola B, Colson P. 2021. Marseilleviruses: an update in 2021. Front Microbiol 12:648731. https://doi.org/10.3389/fmicb.2021.648731
14. Rodrigues RAL, da Silva LCF, Abrahão JS. 2020. Translating the language of giants: translation-related genes as a major contribution of giant viruses to the virosphere. Arch Virol 165:1267–1278. https://doi.org/10.1007/s00705-020-04626-2
15. Abrahão JS, Araújo R, Colson P, La Scola B. 2017. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. PLoS Genet 13:e1006532. https://doi.org/10.1371/journal.pgen.1006532
16. Current ICTV taxonomy release. ICTV. Available from: https://ictv.global/taxonomy. Retrieved 01 Oct 2022.
17. Boughalmi M, Pagnier I, Aherfi S, Colson P, Raoult D, La Scola B. 2013. First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax. Intervirology 56:386–394. https://doi.org/10.1159/000354560
18. Andrade A, Rodrigues RAL, Oliveira GP, Andrade KR, Bonjardim CA, La Scola B, Kroon EG, Abrahão JS. 2017. Filling knowledge gaps for mimivirus entry, uncoating, and morphogenesis. J Virol 91:e01335-17. https://doi.org/10.1128/JVI.01335-17
19. Maruri-Avidal L, Weisberg AS, Moss B. 2013. Association of the vaccinia virus A11 protein with the endoplasmic reticulum and crescent precursors of immature virions. J Virol 87:10195–10206. https://doi.org/10.1128/JVI.01601-13
20. Chatterjee A, Kondabagil K. 2017. Complete genome sequence of Kurlavirus, a novel member of the family *Marseilleviridae* isolated in Mumbai, India. Arch Virol 162:3243–3245. https://doi.org/10.1007/s00705-017-3469-z
21. Blanca L, Christo-Foroux E, Rigou S, Legendre M. 2020. Comparative analysis of the circular and highly asymmetrical *Marseilleviridae* genomes. Viruses 12:1270. https://doi.org/10.3390/v12111270
22. Queiroz VF, Carvalho J, de Souza FG, Lima MT, Santos JD, Rocha KLS, de Oliveira DB, Araújo JP, Ullmann LS, Rodrigues RAL, Abrahão JS. 2023. Analysis of the genomic features and evolutionary history of pithovirus-like isolates reveals two major divergent groups of viruses. J Virol 97:e0041123. https://doi.org/10.1128/jvi.00411-23
23. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. 2021. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLOS Biol 19:e3001430. https://doi.org/10.1371/journal.pbio.3001430
24. Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, Suttle CA. 2023. Taxonomic update for giant viruses in the order Imitervirales (phylum *Nucleocytoviricota*). Arch Virol 168:283. https://doi.org/10.1007/s00705-023-05906-3
25. Doutre G, Arfib B, Rochette P, Claverie J-M, Bonin P, Abergel C. 2015. Complete genome sequence of a new member of the *Marseilleviridae* recovered from the brackish submarine spring in the Cassis Port-Miou Calanque, France. Genome Announc 3:e01148-15. https://doi.org/10.1128/genomeA.01148-15
26. Arantes TS, Rodrigues RAL, Dos Santos Silva LK, Oliveira GP, de Souza HL, Khalil JYB, de Oliveira DB, Torres AA, da Silva LL, Colson P, Kroon EG, da Fonseca FG, Bonjardim CA, La Scola B, Abrahão JS. 2016. The large Marseillevirus explores different entry pathways by forming giant infectious vesicles. J Virol 90:5246–5255. https://doi.org/10.1128/JVI.00177-16
27. Greub G, Raoult D. 2004. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev 17:413–433. https://doi.org/10.1128/CMR.17.2.413-433.2004
28. Greub G, Raoult D. 2002. Crescent bodies of *Parachlamydia acanthamoeba* and its life cycle within *Acanthamoeba polyphaga*: an electron micrograph study. Appl Environ Microbiol 68:3076–3084. https://doi.org/10.1128/AEM.68.6.3076-3084.2002
29. Boratto PVM, Oliveira GP, Machado TB, Andrade A, Baudoin J-P, Klose T, Schulz F, Azza S, Decloquement P, Chabrière E, Colson P, Levasseur A, La Scola B, Abrahão JS. 2020. Yaravirus: a novel 80-nm virus infecting *Acanthamoeba castellanii*. Proc Natl Acad Sci U S A 117:16579–16586. https://doi.org/10.1073/pnas.2001637117
30. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie J-M. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc Natl Acad Sci U S A 111:4274–4279. https://doi.org/10.1073/pnas.1320670111
31. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152
32. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

33. Machado TB, de Aquino ILM, Abrahão JS. 2022. Isolation of giant viruses of *Acanthamoeba castellanii*. Curr Protoc 2:e455. https://doi.org/10.1002/cpz1.455
34. Reed LJ, Muench H. 1938. A simple method of estimating fifty per cent endpoints. Am J Epidemiol 27:493–497. https://doi.org/10.1093/oxfordjournals.aje.a118408
35. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021
37. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using spades de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102
38. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. https://doi.org/10.1093/nar/gki487
39. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119
40. Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol J 6:223. https://doi.org/10.1186/1743-422X-6-223
41. Rodrigues RAL, Queiroz VF, Ghosh J, Dunigan DD, Van Etten JL. 2022. Functional genomic analyses reveal an open pan-genome for the chloroviruses and a potential for genetic innovation in new isolates. J Virol 96:e0136721. https://doi.org/10.1128/JVI.01367-21
42. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704
43. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
44. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
45. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285
46. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096
47. Louca S, Doebeli M. 2018. Efficient comparative phylogenetics on large trees. Bioinformatics 34:1053–1055. https://doi.org/10.1093/bioinformatics/btx701
48. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124. https://doi.org/10.1186/1471-2105-12-124
49. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. 2010. MeV: MultiExperiment viewer, p 267–277. In Ochs MF, Casagrande JT, Davuluri RV (ed), Biomedical informatics for cancer research. Springer US, Boston, MA.
50. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9